



FACULTY OF LAW
Lund University

Emma Schuster

WHEN SOMETHING HAS TO GIVE:
The intersection of Artificial Intelligence, Military
Decision-Making and International Humanitarian Law

JAMM07 Master Thesis

International Human Rights Law
30 higher education credits

Supervisor: Professor Gregor Noll

Term: Spring 2018

TABLE OF CONTENTS

TABLE OF CONTENTS	1
ABSTRACT	3
PREFACE	4
ACKNOWLEDGEMENTS	5
TERMS	6
SECTION I - INTRODUCTION	7
1.1. RESEARCH QUESTION	7
1.2. STRUCTURE OF THE THESIS	7
1.3. AUTONOMY AND MILITARY DECISION-MAKING	9
1.4. THE RELATIONSHIP BETWEEN MILITARY NORMS, IHL, AND AI	9
1.5. THE DOCTRINE OF SUPERIOR RESPONSIBILITY	10
1.6. A MATTER OF URGENCY	11
SECTION II - ARTIFICIAL INTELLIGENCE IN THE MILITARY	12
2.1. THE DUAL-PURPOSE NATURE OF ARTIFICIAL INTELLIGENCE	12
2.2. GOING BEYOND LETHAL AUTONOMY	12
2.3. THE RELATIONSHIP BETWEEN AI AND THE 'ART' OF MDMP	15
SECTION III - CONTEMPORARY PARADIGM: MDMP AND IHL	17
3.1. COMMAND HIERARCHY: THE LEVELS OF WAR	17
3.2. MILITARY LEGAL ADVISORS	19
3.3. THE MANUAL PROCESS OF MILITARY DECISION-MAKING	21
3.4. SUPERIOR RESPONSIBILITY: THE INTERFACE BETWEEN IHL AND WARFARE	24
3.5. CONCLUSION	31
SECTION IV - FUTURE PARADIGM: AI AND MDMP	32
4.1. CADET - A TEST CASE OF A MODULAR AI PLANNING SYSTEM	32
4.2. THE INTEGRATED COURSE OF ACTION CRITIQUING AND ELABORATION SYSTEM	33
4.3. HOW AI WORKS	36
4.4. INCORPORATING AI INTO MILITARY DECISION-MAKING: HOW IT MIGHT WORK	39
4.5. OPERATIONALISATION OF AI DECISION-MAKING INTO THE MILITARY	44
4.6. CONCLUSION	45
SECTION V – FUTURE PARADIGM: AI, MDMP AND IHL	47

5.1. ARTICLE 36 WEAPONS REVIEW	47
5.2. SUPERIOR RESPONSIBILITY: MINIMUM STANDARDS	49
5.3. INDIVIDUAL LIABILITY OF SUPERIORS	53
5.4. INDIVIDUAL LIABILITY OF AI EXPERTS	55
SECTION VI – CONCLUDING REMARKS	56
BIBLIOGRPAHY	58

ABSTRACT

This thesis explores the intersection of three disciplines: military normativity, international humanitarian law, and artificial intelligence, focusing specifically on intelligent systems that assist superiors in conducting the military decision-making process. It explains the extent to which such systems have been operationalised into the U.S. armed forces at the different levels of warfare and discusses the near-future scenario when they might be used for increasingly complex decision-making and planning. It then analyses the effect this will have on the implementation of humanitarian law during the planning process, particularly at the operational level, with an emphasis on the difficulties that arise around the more subjective assessments, such as proportionality and military necessity, and the inscrutability of the cognitive processes of intelligent systems. The analysis underscores the duties of superiors to be aware of the legal implications of their decisions and to retain control over the planning process. Consequently, it argues that the doctrine of superior responsibility is the appropriate legal framework to regulate the use of these systems by establishing international minimum standards. A few examples are given, but the substance of these standards will need to be the subject of further research, requiring interdisciplinary cooperation between lawyers and specialists in artificial intelligence.

PREFACE

This thesis began under the assumption that AI was certainly being used to conduct military decision-making and sought to point out the problems this presents for humanitarian law. However, an investigation into the research and military practices revealed how little operationalisation there has been of this technology into military decision-making, particularly at the higher operational and strategic levels. Further, it became clear that, at least to date, there has been little interest or investment into R&D on this subject by the U.S. military. The possible reasons for this are dealt with in the body of this thesis, but they relate in some part to the traditional attitude of the military establishment towards the complexity of military decision-making. A clear illustration of this attitude came as a result of a communication I had with the author of a piece of research called MATE which used deep learning to create alternative courses of action for tactical commanders. The research was initially funded by DARPA but was ultimately dropped as a result of congressional budget cuts into that area of research (beyond MATE) and the author indicated to me that they showed no interest in picking it up again.

There is little doubt, however, that this attitude will have to change as the technology becomes increasingly sophisticated and the advantages that such an application of AI will provide become clearer, combined with the political pressure on states and other armed groups to remain militarily competitive. The position of the U.S. as the world's military leader (both in terms of its current power and its available resources for research and development) and its interest in remaining so, is the reason for the exclusive focus on U.S. military practices throughout this thesis. The assumption is that where they go, other states will surely follow. In its report on the Future Operating Environment 2035, the U.K. Ministry of Defence notes that the U.S. is likely to retain its position as the world's leading military power to 2035, but that technology will be a key driver in terms of shifts in global power dynamics and that the U.S.' ability to retain its dominant military position will therefore rely heavily on its significant investment into R&D and proposed operationalisation of innovative technologies.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to my supervisor, Professor Gregor Noll, who was generous with his time and ideas, and was able to cut through my jumbled thoughts with razor precision and elucidate exactly what I was trying to say.

I would also like to thank my friends in Lund that made my experience so much richer, particularly Karolina who motivates me to do more and be better.

Finally, my gratitude to Tyron, for always understanding everything.

TERMS

Artificial Intelligence	AI
Course of Action	COA
Course of Action Development and Evaluation Tool	CADET
Defense Advanced Research Program Agency	DARPA
Integrated Course of Action Critiquing and Elaboration System	ICCES
International Committee of the Red Cross	ICRC
International Humanitarian Law	IHL
Judge Advocate	JA
Lethal Autonomous Weapon Systems	LAWS
Military Decision-Making Process	MDMP
Research and Development	R&D
Synchronisation Matrix	SM
United Nations	U.N.
United States of America	U.S.

SECTION I - INTRODUCTION

One of the biggest unknowns about artificial intelligence is what the military applications of autonomous systems will be, and whether and how the international community should approach their regulation. Attempts to address this are being made at various different levels from the United Nations, universities and research institutes to NGOs, civil society organisations and prominent individuals.¹ In recent years, statements from politicians regarding AI being the next arms race have captured the imagination of the media, and have fueled the interest of researchers and lawyers, amongst others. It may yet prove to be true. But it is important first to distinguish between what is technologically possible, what is being researched and tested, and what the potential of AI holds for the future. A proper understanding of the state of the technology will enable lawmakers to recognise the advantages AI will bring to militaries and understand the accompanying challenges, in order to be able to address the latter while allowing for the former.

1.1. Research question

The question this thesis seeks to answer, is as follows:

How would an artificially intelligent system that assists with military decision-making, affect the implementation of IHL at the tactical, operational and strategic levels of war?

In order to answer that, it looks at two additional questions, namely:

- (i) What is the state of the technology in terms of being able to conduct decision-making and planning, and what is the extent of use at the various levels of the military?
- (ii) Which tools or mechanisms of IHL exist that can address the changing relationship between human and machine in respect of military decision-making and planning, and ensure that implementation of its rules and principles is not negatively affected?

1.2. Structure of the Thesis

The remainder of this thesis is divided as follows: This section I provides an introduction to some of the concepts and concerns underlying the thesis.

Section II will locate the topic of the thesis within the broader debate about AI in the military.

Section III will describe the current military and legal positions on military decision making, beginning with an explanation of the levels of warfare according to military theory and then summarising the existing procedure of military decision-making and the role played by legal advisors according to U.S. military doctrine. Next the doctrine of superior responsibility is introduced. This section is not intended to be a full primer on IHL, but rather to outline the most critical assumptions underlying the law of superior responsibility that will be pertinent to the future paradigm of military decision-making in the age of artificial intelligence.

Section IV is the discussion of this future paradigm. It begins with an experimental piece of research that shows how basic AI can be applied to the military decision-making process, primarily at the tactical and operational levels. After summarising some key AI techniques, it then introduces what

¹ See for example *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, (2015) available at <https://futureoflife.org/open-letter-autonomous-weapons#signatories>; and *An Open letter to the United Nations Convention on Certain Conventional Weapons*, (2017) available at <https://futureoflife.org/autonomous-weapons-open-letter-2017/>

is novel about this research, which is (i) an analysis (with some unavoidable speculation) of how AI will affect military decision-making in the future, together with a discussion about the benefits of the technology to operational success, and the challenges it poses to the implementation of IHL during warfare and (ii) a description of the state of usage of this technology at the different levels of military decision-making, and possible explanations for it, linking the scenarios in which AI technology works best to the functions of the three levels of warfare.

Section V brings the current and future paradigms together in an examination of how this future will affect the implementation of IHL, and how IHL is prepared to address it by relying on the doctrine of superior responsibility to set minimum standards applicable to superiors during the prosecution of war. Linking this to the broader scheme of IHL, which in general struggles to hold individuals accountable except in limited circumstances for specific violations and the worst war crimes, it emphasises instead the reliance on the principle of responsible command.

Finally, Section VI draws together the three strands; military, legal, and AI, in an answer to the research questions posed above.

1.3. Autonomy and Military Decision-Making

In the report of the CCW conference in 2017,² Brigadier Bezombes describes four levels of automated systems with increasing levels of autonomy that have military application. The first three, teleoperated, supervised and semiautonomous systems, have been operationalised and are in use by modern militaries in some form or other.³ The final level represents the promise of AI for the future, namely fully autonomous, unsupervised systems that are able to make and execute decisions independently of humans. Although still only a budding possibility, R&D is taking place at a rapid pace, particularly in civilian industries⁴ which are involved in their own kind of commercial ‘arms’ race. So while the experts say that the reality of autonomous weapon systems is some way off, the reality of humans and machines sharing tasks that were previously thought to be exclusively human is already present. To date the balance of task-sharing in military decision-making remains weighted in favour of humans but, with recent breakthroughs in the application of machine learning techniques, this balance appears to be shifting. High frequency trading platforms, for example, use machine learning to adapt and react to financial markets in nanosecond intervals. Humans are unable to follow these processes in real time, to say nothing of intervening in them, and so the balance of task-sharing is shifted to the AI. AlphaGo is an AI that can determine millions of potential actions and adversary reactions in the context of the game of GO, again at speeds and volumes that humans cannot hope to match.⁵ One can easily imagine the lateral application of this technology into the military domain, to strategic and operational planning and military decision-making in particular.

Military decision-making is, in fact, well suited to human-AI task-sharing. Currently it is a largely human driven process, but the introduction of an intelligent system that assists, and potentially ultimately replaces the human planner, would bring many advantages to the process on the strategic, operational, and tactical levels of warfare. These advantages may even extend to the better implementation of humanitarian principles, provided the use of such systems is managed carefully, timeously and according to a framework of IHL.

1.4. The relationship between military norms, IHL, and AI

In spite of these advantages, and the technological advances taking place, thus far there has not been an explosion of applications of AI decision-making technology in a military context, particularly at higher levels of the command structure. There are several factors to explain this, including technical and institutional, but influential voices from within the U.S. military community have been calling for increased automation of the military’s planning processes since the beginning of this century,⁶ and have stressed the need to “*create fast new planning processes that establish a new division of labor between man and machine. [...] Decision aids will quickly offer suggestions and test alternative courses of actions.*”⁷ The emphasis on automation of decision-making raises a number of questions from a regulatory perspective, particularly regarding the implementation of IHL. Will these systems be computationally able to take account of law? How will they affect the implementation of IHL during decision-making and planning at each echelon of warfare? In hard cases involving war crimes and serious breaches of IHL, how will the law of individual responsibility be affected, given the inscrutability of AI? The answers to these questions will determine which aspects of the process can and should be handed over to intelligent systems, and which must remain under human control.

² Geneva, *Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, (2017), CCW/GGE.1/2017/CRP.1

³ For example, the autonomous reaper and predator drones that are used for engagement

⁴ Cummings, M., *Artificial Intelligence and the Future of Warfare*, (2017) Chatham House, p. 9

⁵ DeepMind, *AlphaGo: The Story of AlphaGo so far*, available at <https://deepmind.com/research/alphago>

⁶ Ground, L., et al., *Coalition-based Planning of Military Operations: Adversarial Reasoning Algorithms in an Integrated Decision Aid*, (2015), CoRR

⁷ *ibid*, p. 1

Some of this work will require technical responses from AI experts, computer scientists and engineers, which illustrates the demand for interdisciplinary work on this topic. Clarity regarding how susceptible and obedient AI systems will be to human oversight will help the legal community to decide whether and how to regulate the development and use of the technology. The legal and ethical parameters need to inform the development of the technology, while at the same time the technological advances will dictate the direction of the regulatory responses. However, there are some parts of these questions that can be addressed in the meantime by looking at the way IHL currently governs military decision-making, and deciding whether it has the framework, and indeed the mandate, to regulate the use of intelligent systems.

What is clear is that the introduction of AI into the hitherto predominantly human exercise of military decision-making brings together three distinct disciplines, namely the military, legal and AI, each with its own goals, ontology and internal logic. Sometimes these logics converge and other times they come into stark contradiction with each other. The intersection of military normativity, IHL and AI will inevitably result in a cost imposed on the logic of one or the other as compromises are made, conceptually and in the field of battle. This contest has played out before. The very existence of IHL is at once a recognition of the reality of warfare as a tool of human relations and a constraint on states' rights to use force to obtain victory at all costs, which are in constant tension both at the level of legislation and on the battlefield. The acknowledgement that the nature of war is chaotic, violent and politically driven, is the basis of the pragmatic balancing of military necessity with the principle of humanity that underlies IHL.⁸

IHL seeks to humanise war by minimising the harm it causes, through various mechanisms. It seeks first and foremost to preclude violations through preventative measures. Where prevention fails, the question will then turn upon the nature of the breach and the person involved: is it a war crime or other grave breach, a serious violation or a minor transgression? Does it involve a systematic abuse of power, or the rogue actions of a few individuals? Depending on the answer to these questions there may be either individual or state liability. IHL does not have mechanisms of accountability for minor breaches, but it attempts to tackle the culture of impunity that regularly accompanies wartime atrocities, particularly in respect of high-ranking military and political officials, by imposing criminal responsibility either directly through international criminal law, or indirectly via the doctrine of superior responsibility.⁹

1.5. The doctrine of superior responsibility

With the disruptive and possibly transformative effect that AI may have on warfare comes the potential to radically transform the strategic balance which currently determines how militaries function, including the process of decision-making,¹⁰ and the current lines of accountability under IHL. Their ability to outperform humans in speed and quality of decision-making raises the question of what the role for the human decision-maker (and legal advisor) is in the process of MDMP. Furthermore, the already complex and opaque process of military decision-making is covered by an additional shroud of impenetrability. International lawmakers must therefore be aware of how these tools will affect the implementation of IHL norms in warfare, and how they might address this. One way is on the basis of the doctrine of superior responsibility.

The doctrine of superior responsibility is primarily concerned with holding superiors accountable for crimes committed by their subordinates if they fail to prevent the crime or to adequately punish the perpetrators. However, the doctrine has another goal which is to set internationally sanctioned minimum standards of conduct for superiors in the prosecution of war. These two aspects of the

⁸ Kolb, R., *Advanced Introduction to International Humanitarian Law*, (2014) Edward Elgar, p. 75-76

⁹ The term can be used interchangeably with command responsibility. 'Superior responsibility' is preferred as it encompasses both international and non-international armed conflicts, but 'command responsibility' is used where the context requires

¹⁰ Ayoub, K., and Payne, K., *Strategy in the Age of Artificial Intelligence*, (2016), JSS, p. 806

doctrine at times seem to be in competition with one another, and consequently give rise to some debate about the details of the doctrine including its scope, the required standard of *mens rea* of the accused, and the nature of the fault attributable to superiors.¹¹ Despite this friction, however, the standard-setting function is clearly an important aspect of the doctrine and, what is more, is not a completely new idea. The two-pronged approach resembles the structure of IHL more generally in that it too pursues its implementation largely through parallel mechanisms. In the case of IHL these are preventative measures, control, and suppression particularly in the form of direct sanctions. Given that the latter two mechanisms are problematic in many respects, IHL relies largely on the voluntary, spontaneous compliance of states and other actors.¹² This softer form of implementation plays an invaluable role in realising the ambition of IHL, namely the humanisation of warfare. In the future of increasingly automated warfare, not only in respect of lethal weapons but also decision-making and planning, it is possible that increased reliance on preventative measures will gain in importance even more. It is, therefore, a branch of IHL that needs substantially more consideration in light of the proposed challenges that AI presents to the distribution of decision-making tasks between humans and machines.

1.6. A matter of urgency

AI poses a unique challenge in that, by its nature, it may soon lead to a situation whereby humans will be unable to regulate it retroactively. Once the technology is sufficiently advanced, there is the possibility that AI systems will become more intelligent than humans, and no longer entirely subject to our will or control and it is this threat that motivates the legal community, among others, to try to grapple with issues that are not yet fully understood, in an effort to control the development of the technology itself. This work must be done timeously and continuously. While the nature of war is immutable, its character is ever changing and developing according to technological advances and changes in cultural mores and it is the critical task of the international lawyer in the face of a new weapon or military activity to establish how existing law applies and with what effect.¹³

¹¹ Mettraux, G., *Command Responsibility*, (2009), Oxford University Press, p. 20

¹² Kolb (n. 8) pp. 187-197

¹³ Doswald-Beck, L., *Some thoughts on the Computer Network Attack and the International Law of Armed Conflict*, (2002), Naval War College, p. 163

SECTION II - ARTIFICIAL INTELLIGENCE IN THE MILITARY

To begin with, because it is a contested concept, it must be established what AI means in this thesis. For some, it refers to any computer system that mimics human intelligence of even the most basic form. On the other end of the spectrum are those who consider AI to refer only to complex systems that can conduct the most meta-level decision making. Here however, it refers to any computer systems that can recognise a scenario ('if') and respond either with action or recommendation ('then').

2.1. The dual-purpose nature of artificial intelligence

There are some challenges and critiques that are unique to artificial intelligence, and some that are true of any emerging technology. One example of a general technological challenge is the dual-purpose nature of AI and its potential for abuse. Some AI algorithms, for example, are being used for planning disaster response operations. A report from Chatham House on AI states that

“in the coming years, these algorithms are likely to improve, and as they are able to undergo further modification, such as through genetic evolution, learning and adaptation, their abilities will sharpen. Indeed, there is no ostensible reason why the primary objectives of the UN’s Department of Field Support could not be met by using planning AI, thereby automating the majority of its tasks. By doing this, time spent on forecasting and planning would fall, and its implementation would cut costs related to human resources and training and remove many redundancies and barriers in supply chain logistics. This would improve the efficiency of any service centres and potentially result in rapid deployment of forces at a reduced cost.”

The report goes on to say that

“Governments, NGOs and civil society groups can also avail themselves of this AI. NGOs and civil society groups may in fact be best placed to trial these technologies, as they are often not hampered by political obstacles or byzantine bureaucratic rules. They may have greater flexibility to try out new approaches and improve confidence in those ideas. This would help immensely in various human security-related situations, such as complex humanitarian crises – those that are combinations of political and natural disasters – as they could examine vast amounts of data relating to available resources, use satellite imagery or images from surveillance aircraft to map affected terrain, as well as find survivors, and thereby estimate the necessary resource requirements given limitations on time, goods and manpower.”¹⁴

The application suggested in this report is ostensibly beneficial, but subject to vastly different uses under changing conditions and less benign powers. The capabilities of AI can be calibrated differently so that the same techniques could tackle minor concerns for truly constructive purposes or cause widespread devastation.¹⁵ The tension between the beneficial potential of these systems and the possibility of abuse is precisely why it is necessary to determine how to ensure clear lines are drawn regarding what constitutes abuse. It may not always be used for sinister, genocidal intent but may also mask non-compliance with the more pedestrian restrictive provisions of IHL. There is often resistance to the idea of IHL within the military domain, as a result of the distinct internal logic it represents that at times contradicts pure military reasoning.

2.2. Going beyond Lethal Autonomy

How the regulation of AI in the military is addressed depends very much on the parameters that are adopted to constrain the problem. In order to navigate the myriad of competing interests, the debate

¹⁴ Roff, H. *Advancing Human Security through Artificial Intelligence* (2017), Chatham House, p. 8

¹⁵ Ayoub and Payne (n. 10) p. 810

within traditional international legal structures so far has been narrowly construed to address artificial intelligence only insofar as it is manifested as lethal autonomous weapon systems (LAWS)¹⁶ or lethal autonomous robotics.¹⁷ It is easy to understand why an organisation such as the United Nations would limit itself to this narrow definition as it is more manageable, does not impinge on the various political and industry wide pressures, and is a relatively straightforward path to reaching an agreement. Thus a growing consensus has emerged that LAWS constitutes a significant threat to the legal and ethical framework underlying the laws of armed conflict.¹⁸ For now, the international community has come up with the notion of ‘meaningful human control’ in an attempt to establish a universal standard - a cut-off point - after which the development and use of LAWS will in all likelihood be banned, or at least subject to intense scrutiny.

The U.S. has also been developing policies on the development and use of semi-autonomous and autonomous weapons.¹⁹ An autonomous weapon system, according to DoD directive 3000.09 is defined as one that “*once activated, can select and engage targets without further intervention by a human operator.*” Although this definition also includes “*human-supervised autonomous weapons systems that are designed to allow human operators to override operation of the weapon system*” it is still clearly intended to refer to those systems that function as weapons in the traditional sense of engaging directly with targets.

However, in spite of this narrow construction of the issue, the potential applications of intelligent systems to the military go far beyond LAWS, including systems that are not directly lethal in themselves, but that nevertheless will have consequences for IHL. Some examples of the use (current and/or potential) of AI systems other than LAWS are (i) in battlefield simulations for training; (ii) performing intelligence, surveillance and reconnaissance functions (iii) interpreting and analysing huge amounts of battlefield data and (iv) assisting with the production and testing of battle plans. It is conceivable that in the future, artificial intelligence will allow these functions to communicate directly with one another, so that battlefield information is gathered and interpreted, various COAs designed and tested, and directly communicated to a lethal weapon system, without the need for human input. As much as the law may require human involvement in the process (so that there is someone to point to when things go wrong), such involvement will fall short of being meaningful if the individual is not privy to the cognitive process having taken place in the interim steps before execution of a lethal action. The danger of focusing only on a limited definition of ‘lethal autonomy’ is that other applications, and their effects on the implementation of IHL are neglected.

It is critical to remember that the goal of artificial intelligence is to resemble, and ultimately surpass general human intelligence. Human intelligence works not in discrete pockets of capability but rather as a complex, continuous process. At its best, a human brain is able to apply knowledge, experience and abilities to any problem it is presented with across all skills and subject matters. To understand the true potential (and threat) of AI, it must be seen as a similarly continuous system, and not only as applications of the technology to isolated competencies. On the whole, AI is currently only being applied to separate and distinct tasks that that have previously been conducted either wholly, or principally, by humans. In the military context this spans intelligence gathering, through the planning process, to the execution of operations. Yet it is the nature of technology generally that it is an inevitable, incremental forward march, and automation is no different.²⁰ With

¹⁶ *Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems* (n. 2)

¹⁷ Heyns, C., *Report of the Special Rapporteur on extrajudicial summary or arbitrary executions*, (2013), United Nations General Assembly, A/HRC/23/47

¹⁸ Acheson, R., (ed.), *Losing Control: The Challenge of Autonomous Weapons for Laws, Ethics, and Humanity*, CCW Report on the CCW Group of Governmental Experts on Lethal Autonomous Weapons Systems (2017), WILPF

¹⁹ Saxon, D., *A Human Touch: Autonomous Weapons, Directive 3000.09, and the “Appropriate Levels of Human Judgment over the Use of Force*, (2016) Cambridge University Press

²⁰ Anderson, K., and Waxman, M., *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won’t Work and How the Laws of War Can*, (2013), p 4.

the Internet of Things, AI has the potential for these previously separate functions to merge into smooth lines of communication, forming a continuous super-system. The more these systems communicate directly with one another, the less the human has to say to the outcomes of decision-making, and it is this potential that threatens the concept of effective human control. There is the possibility that the systems are built so as to allow for, possibly even require, ultimate human control, but it is not guaranteed either that this will be technologically feasible nor that developers and users will abide by it. Furthermore, AI decision-making systems illustrate how the challenges of AI extend beyond just lethal autonomy. Military decision-making is part of the greater process that ultimately affects how a particular armed conflict, or specific engagement, is conducted and AI has the potential to confer benefits at any level, from minor tactical advantages to overall strategic advantages.

A better understanding of the technology will help to clarify why it poses a unique challenge to IHL and why the definition needs to be expanded to review non-lethal systems too. There is, therefore, a pressing need for increased communication between the lawmakers, ethicists, computer scientists and engineers. Without a proper understanding of what has been achieved in laboratory settings, actually operationalised, and what is possible to achieve in the future, it is difficult for lawmakers to determine what constitutes a genuine threat versus mere conjecture, and what is needed to ensure that the technology is developed responsibly. Traditional international legal institutions have taken a limited role thus far (besides its engagement with LAWS), resulting in a bottom-up process of lawmaking, not only in respect of technical standards but ethical, moral and political concerns too.²¹ As a result, efforts to deal with ethical concerns are taking the form of ‘self-regulation’ from within the AI industry; from scientists, think-tanks and universities. While these are valuable contributions, international lawmakers must also be prepared to adapt to the technological advances in AI timeously. In the face of ubiquitous AI, retrospective regulation may prove difficult, not least because customary international law will be formed by state action if it is left unregulated.

The need for a broader definition to address the consequence of AI systems for IHL is therefore imperative. The Harvard Law School introduced the concept of ‘war algorithms’ in 2016, which it defines as “*any algorithm that is expressed in computer code, that is effectuated through a constructed system, and that is capable of operating in relation to armed conflict*”²². This kind of expansive approach is better suited to the nature of AI and as it is more adaptable as the technology moves towards a continuous system, rather than focusing on one area of use of the technology. In going beyond LAWS, it encompasses more than just the question of whether or not machines can and should be trusted to make life and death decisions. By putting algorithmically-derived choices and decisions at the centre of discussion regarding autonomy systems in war²³, it takes in any AI system that is capable of making and effectuating a decision or choice that would ordinarily require the application of human intelligence. Finally, it opens up the discussion of legal responsibility for wrong-doing under international criminal law, which depends on proving the mental state of the accused. This already opaque process is further complicated by the use of AI systems to aid military decision-making.

In the meantime, the technology continues to advance in diverse spheres of civilian life including transport, medicine and law enforcement, which international lawmakers have no authority to arrest. By having a better understanding of how the technology works, policy and law makers can influence how they are designed. Among computer scientists and engineers, the approach is predominantly to push for as much automation as possible. It is for policy and law makers to ensure that they work within a set of normative rules and clear design criteria.²⁴

²¹ Burri, T., *International Law and Artificial Intelligence*, (2017), German Yearbook of International Law (2017/18), p. 9

²² Lewis, D., Blum G., and Modirzadeh, N., *War-Algorithm Accountability*, (2016), HLS PILAC, p. 10

²³ *ibid*, p ii

²⁴ Cummings, M., *Man versus Machine or Man + Machine?* (2014) IEEE Intelligent Systems, p. 62

2.3. The relationship between AI and the ‘art’ of MDMP

In “On War”, Carl von Clausewitz describes the process of decision-making by relating knowledge to art. He argues that “all thought is art” and that in the human mind, art and knowledge are inseparable. Clausewitz reasons that art begins at the point where a human being draws ideas from perceptions to be used in judgment. He goes on to say that the act of perception, including sensory perception, is itself a form of judgment, and so is also an art; “*it is impossible to imagine a human being capable of perception but not of judgment or vice versa*” and therefore knowledge, taken from perception, and art are inextricably linked.²⁵ In respect of military commanders, Clausewitz writes that a commander need not be a master of history, human nature or military skills, but must have sufficient knowledge about each of these to be able to form sound judgments and that “*this type of knowledge cannot be forcibly produced by an apparatus of scientific formulas and mechanics; it can only be gained through a talent for judgment, and by the application of accurate judgment to the observation of man and matter.*”²⁶ Throughout history, new technologies have disrupted the character of war over and over, but none has presented so significant a challenge to this notion of military decision-making as an exclusively human exercise until the introduction of AI, especially insofar as it has the potential to share the process of decision-making, and possibly ultimately replace it. Sceptics of AI continue to argue that there is no replacement for the art of human judgment.

However, discussions about AI have invoked questions about what human intelligence entails, whether any thought is truly creative and original, and even if it is, whether it is possible to emulate the processes of the human brain in a manner that would enable a machine to have original creative thoughts. Modular AI is becoming increasingly successful at the first part of Clausewitz equation, namely its ability to perceive its surroundings using a range of sensory applications, and to record and interpret the information. Effectively that means it can take over many of the tactical decisions previously made by low ranking officers and soldiers. Yet in Clausewitz’ analysis, perception and judgment are bound up as two parts of one process. So, while AI is able to ‘perceive’ in a way similar to humans, it is the understanding and interpretation of that perception, that constitutes judgment. There are different ways of understanding what happens to ‘judgement’ in this context. The first is that the ontology of the AI that perceives the situation at the tactical level includes an ability to judge its own perception, which would push traditionally strategic thinking into a new domain, one that is closer to the tactical level. Alternatively, it may be that the act of perception is not connected at all to an act of judgment at the tactical level, but instead is transported to the higher operational and strategic levels. The interconnected process of perception and judgement is then acted out there. Likely it will be a combination of both, depending on the complexity of the particular scenario and the sophistication of the AI.

At any rate, current strategic and operational MDMP requires an ability to understand the battlefield information regarding the assets and capabilities of friendly and enemy forces that must be selected, filtered, and analysed. The quality of the assessment will depend on many factors including the experience and competency of the planning staff, time and resource constraints, the nature of the mission and the adversary, the preferences of the commander, and the accessibility and quality of information. There has always been a ‘fog’ in this regard in that the environment is difficult to fully comprehend given that the commander relies on others to collect and communicate all relevant information, and that it often takes place under hostile conditions. The fog of war, from Clausewitz to today, largely resulted from a lack of good information about the battlefield and is one reason that international law has difficulty in holding individuals accountable for breaches of IHL. AI offers a way to achieve a much improved understanding of the battlefield in its ability to collect vastly more information and conduct a better quality analysis. However, the amount of information it is able to obtain gives rise to a new problem: the incapacity of humans to view, analyse and interpret it without further help from automated (or autonomous) systems. The ever-changing character of war and the complexity of the new battlefield, together with a rise in the technology and availability of big data,

²⁵ Clausewitz, C., (1989) p. 148

²⁶ Clausewitz, C., (1989) p. 146

means that militaries will inevitably have to rely on intelligent tools to assist them with the planning of operations in one way or another. But an overreliance on these systems combined with their black-box cognitive processes, introduces a new fog of war, the AI fog. A big question for the future, then, will be whether the use of these tools increases the fog of war, or helps to clear it up by providing a more accurate picture of the battlefield and enabling decision-makers to explore and analyse a greater number of potential COAs in much less time and based on better information.

SECTION III - CONTEMPORARY PARADIGM: MDMP AND IHL

Every formal modern military has its own doctrinal method for generating and monitoring decisions in military operations. While each doctrine may differ in its particulars, they will have at least these fundamental features in common: the process is triggered by receipt of a mission from a higher authority, which is analysed and broken down into to a set of tasks. The information is then used to develop one or more potential courses of actions (COAs), which are studied and compared against one another and potential enemy COAs – using a set of predetermined criteria – to determine the best COA. The COA is then executed and monitored and evaluated to ensure that it converges with the commander’s original intent for the mission.

What follows is a closer look at the military decision-making process of the United States armed forces (MDMP), and the function of legal advisors in the process, according to U.S. military manuals. The formal MDMP described below applies to any operation that engages the U.S. armed forces whether at the strategic, operational, or tactical level, and whether deliberate or time-sensitive planning, with the appropriate adaptation depending on which iteration is at work.²⁷

3.1. Command hierarchy: The levels of war

Discipline and accountability are crucial principles to the successful functioning of armed forces, whether formal militaries or informal militias. Militaries are deliberately structured to ensure that decisions are made at the appropriate level for operations to be conducted as successfully as possible, which means that those required to take decisions are adequately qualified, but also that those who carry out orders are not concerned about whether a decision is a good one or not. The only time when a soldier, for example, is required to act outside of his or her orders is when they are manifestly unlawful.²⁸ Military structure, therefore, is tied closely with military discipline and accountability,²⁹ and the higher up the chain of command it goes, so the nature of responsibility changes accordingly.

Clausewitz describes the art of war in a wide and a narrow sense. In its wider sense, it encompasses all activities that exist for the sake of war, including raising, arming and equipping and training the fighting forces. In its narrow sense, however, what he calls the conduct of war, it refers to “*the art of using the given means in combat.*”³⁰ By this, Clausewitz is referring to the planning and the conduct of hostilities. Based on the recognition that hostilities take place over a collection of single and distinct acts (engagements), he draws a further distinction between the planning and execution of each engagement on the one hand and coordinating the overall conduct of fighting towards furthering the object of the war on the other. The former is what he calls tactics, the latter strategy. Tactics, then, is how the armed forces are used in an engagement and the form of each engagement, while strategy refers to how engagements are used to achieve the overall objective of the war.³¹ Although these two undertakings are considered separate, they interact with one another throughout a conflict, being temporally and spatially intertwined. Strategic and tactical decisions may cover the same empirical categories,³² but the nature of the activity will depend on whether it relates to the form of the engagement, in which case it is tactical, or to the significance of the engagement, when

²⁷ This process can become complicated over time in respect of joint operations involving different branches of the armed forces. However, the basic MDMP formula remains the same. There are separate manuals relating to coalition military operations that seek to integrate the U.S. MDMP with the doctrines applying to friendly forces.

²⁸ See for example ICRC, Customary IHL Rule 154 which states that “*Every Combatant has a duty to disobey a manifestly unlawful order.*”

²⁹ U.S. Army Regulation 600-20 *Army Command Policy* Chapter 4 *Military Discipline and Conduct*

³⁰ Clausewitz, C., *On War*, (1989), Princeton University Press, p. 128

³¹ *ibid*

³² Clausewitz uses marches, camps and billets as examples, but a contemporary example may be the cooperation of different functions with the armed forces; the army, navy and airforce)

it is strategic.³³ Clausewitz' categories have not altered much in the intervening years, and continue to play an important part in military theory today, with the exception of the addition of a bridging category that connects the tactical to the strategic in the theatres of war. According to U.S. modern military theory, war is divided into three levels which correspond to the divisions in the planning hierarchy,³⁴ strategic, operational and tactical,³⁵ but the doctrinal separation between the objectives behind the war and the means that are used to achieve them is maintained.

Strategic level

States (and sometimes non-state actors) pursue their national interests through various means including through economic, political, and military instruments – not necessarily in that order. Military strategy is thus a component of national strategy pertaining to the development and use of states' military resources, which can be wielded to threaten or use force to ensure the states' security objectives. It involves setting the conditions of conflict, ascertaining the theatres of war, allocating military forces and assigning broad goals to each warfighting function.³⁶ In the U.S. this emanates from the Joint Chiefs of Staff although strategic decision-making also involves other high-level commanders that are responsible for determining the strategic objectives of the armed forces.

Operational level

The operational level was not present in Clausewitz' explanation but entered mainstream military studies and the vocabulary of the U.S. armed forces in the early 1980s as a bridge between the strategic and tactical levels. The operational level comprises those military forces employed to accomplish the strategic goals in the theatre of operations through planning and conducting military campaigns and major operations.³⁷ Commanders at the operational level establish separate objectives linking the strategic objectives to tactics, and then sequence events, initiate actions, and apply military resources towards achieving those operational objectives.

However, according to U.S. Army doctrine, no echelon of the command hierarchy is solely concerned with the operational level. Key personnel at the operational level are the army, army group and joint force commanders. These commanders are aware of the other instruments of power, but themselves are responsible primarily for military concerns, which they accomplish through planning and executing operations. The distinction between strategic and operational level is helpful conceptually but also somewhat arbitrary as either one could be carried out under one commander without having to impose an artificial divide.³⁸ Furthermore, in modern warfare this role is increasingly being filled by tactical commanders.

Tactical level

Campaigns and battles are carried out on the tactical level, with the emphasis on the application of military force to achieve successful outcomes of engagements.³⁹ Combat objectives are assigned to tactical units from the operation level, and battles and engagements are then planned and executed to achieve these objectives. Commanders at the tactical level are focused on purely military matters, namely conducting successful engagements by arranging and manoeuvring of various elements of combat in relation to each other and to the enemy.⁴⁰

³³ Clausewitz (n. 30) p. 132

³⁴ U.S. Army Field Manual 100-5, *Operations*, (1993)

³⁵ *ibid*; also see Lt. Col. Newell C., *The Levels of War*, Army, (1988), 26-29

³⁶ FM 100-5 (n. 34) pp. 9-11

³⁷ *ibid*; Jones, M., *The Operational Level of War: A Primer*, (1988) U.S. Army War College, p. 1

³⁸ Jones (n. 37) pp. 6-7

³⁹ FM 100-5 (n. 34) pp. 9-11

⁴⁰ Jones (n. 37) p. 6

The chain of command, therefore, serves a few functions which all aim at ensuring the best possible military result. It enables fast, clear and efficient transmission of orders between seniors and subordinates (with minimal opportunity for wasted time and misinterpretation), which in turn allows commanders at each level to achieve their primary function. It also secures that each command echelon is given the proper degree of responsibility to accomplish its task, and to have the correct personnel held responsible for their activities with overall responsibility falling to the commander.⁴¹ As, however, that the division is not always strict, particularly concerning modern ‘block warfare’ whereby traditionally low ranking, small-unit tactical commanders are required to make what are essentially operational and even strategic decisions.⁴² As Clausewitz wrote, it is not the location or even the empirical category in question, but the nature of the decision that characterises it as strategic or tactical.⁴³

3.2. Military legal advisors

Article 82 of Additional Protocol I requires all contracting parties to ensure that legal advisors are available, when necessary, to advise military commanders at the appropriate level, on the application of the law and the appropriate instruction to be given to the armed forces.⁴⁴ The wording is not prescriptive regarding the content of the role of the legal advisor besides being primarily preventative and operational, and most states, therefore, pass their own rules to ensure compliance with this requirement and to provide detail and clarity on this role. The legal advisor is concerned with preparation, planning and training purposes during peacetime, as well as legal issues that arise during a conflict. States must also specify the levels at which legal advisors are required to provide expertise, which will differ depending on the type of service (advisory or training) and the nature of the decision-making.

There are, of course, those ideologically resistant to an increased role of military advisors in the conduct of military operations, with some members of the armed forces believing that law has come to create more problems for warfare, rather than any kind of solution. Professor Betts regrets the role of the law and lawyers in the Balkan War. He holds that lawyers “*constrained even the preparations for decisive combat*” and that

“one of the most striking features of the Kosovo campaign, in fact, was the remarkably direct role lawyers played in managing combat operations – to a degree unprecedented in previous wars ... The role played by lawyers in this war should also be sobering – indeed alarming – for devotees of power politics who denigrate the impact of the law on international conflict ... NATO’s lawyers ... became in effect, its tactical commanders.”⁴⁵

Other commentators see the legal advisor not as a threat, but as another member of the staff, providing guidance and information that contributes to the overall set of factors the commander takes into account, but ultimately no more. “*This is the critical point: military lawyers advising national leaders and commanders are not decision-makers. They are counselors to their commanders.*”⁴⁶ The ICRC Advisory Service confirms that the commanders always retain a position of leadership and responsibility in the decision-making process. The legal advisor is limited to briefing senior officers in increasingly complex legal environments.⁴⁷

⁴¹ AR 600-2 (n. 29) p. 6

⁴² Dunnaback, J., *Command Responsibility: A small-unit leader’s perspective*, (2014) Northwestern University Law Review, Vol. 108 No. 4 1385 at p. 1411

⁴³ Clausewitz (n. 30) p. 130-131

⁴⁴ Also captured in ICRC Customary IHL Rule 141

⁴⁵ Betts, R., *Compromised Command*, Foreign Affairs, (2001), at 126 (Review of Wesley Clark’s, *Waging Modern War: Bosnia, Kosovo, and the Future of Combat*)

⁴⁶ Lohr, M. and Gallotta, S., *Legal Support in War: The Role of Military Lawyers*, (2003) Chicago Journal of International Law, p. 468

⁴⁷ ICRC, *Legal Advisers in the Armed Forces*, (2003) Advisory Service on International Humanitarian Law, p. 2

Military advisors in the U.S.

Brigadier General Charles Dunlap said, “*savvy American commanders seldom go to war without their attorneys.*”⁴⁸ This sentiment is reflected in the formal role of legal advisors in the U.S. military decision-making process. MDMP is used for all kinds of operations from large-scale, complex strategic manoeuvres down to individual targeting operations and the steps of the process are in principle the same whether a high-level strategic commander or a field commander in charge of only a few people is conducting it. The size and composition of the planning staff, including the prevalence of the legal advisors, will depend on the nature of the operation and the level at which the process is being conducted. However, it is evident from the U.S. Army Operational Law Handbook, that the role of a military lawyer at the division level or higher, at least, is to work closely and on a regular basis with the planning staff.

The U.S. program dedicated to ensuring that all military operations conducted by the U.S. comply with IHL and that measures are taken to prevent violations, requires commanders to “[e]nsure that qualified legal advisers are immediately available at all levels of command to provide advice about law of war compliance during planning and execution of exercises and operations”.⁴⁹

So to provide the necessary support, judge advocates (JAs)⁵⁰ participate in various aspects of the commander and staffs decision-making. During mission analysis they are required to identify and assess potential legal issues, communicate up and down the chain of command, and review plans and orders. Fundamental to all this is achieving a thorough situational awareness,⁵¹ which includes the pre-battle planning process as well as planning for actions during the conduct of hostilities. A Joint Publication from the DoD for guidance on joint operations, for example, requires JAs to be involved in targeting operations

“Due to complexity and extent of international law considerations involved in the joint targeting process, a JA must be immediately available at all levels of command to provide advice about law of war compliance during planning and execution of exercises and operations. Early involvement by the JA will improve the targeting process and can prevent possible violations of international or domestic law.”⁵²

They are also required to review all the internal policies, rules of engagement, directives and operational plans to ensure their consistency with the international obligations of the U.S., requiring them to be in the command centre. It is the extended responsibility of the military lawyer not only to prosecute war crimes but to ensure compliance with the laws of armed conflict in operations that has led this role of advisor to the commander, including providing support in the planning of operations.⁵³

There are two things to note about legal advisors before looking at the process of MDMP. The first is that the role is increasingly important up the command chain. Accordingly, those operating at a strategic level have much more contact with legal advisors than a small unit commander making tactical, ‘do-or-die’ decisions on the ground, which is sensible given the increasing complexity of the decisions, and the degree of responsibility of strategic and operational superiors. As a result of the decentralisation of modern warfare, tactical commanders are being accompanied by legal advisors more and more, but the principle remains the same: decisions of a strategic or operational

⁴⁸ Col. Dunlap, C., *Law and Military Interventions: Preserving Humanitarian Values in 21st Century Conflicts* (2001), p. 6

⁴⁹ U.S. DoD Directive 5100.77 at 5.3.3. in Lohr and Gallotta (n. 46) p. 470

⁵⁰ Judge advocate is the U.S. term for a military lawyer or legal advisor

⁵¹ Lohr and Gallotta (n. 46) p. 473

⁵² Lohr and Gallotta (n. 46) p. 472

⁵³ Lohr and Gallotta (n. 46) p. 471

nature engage more legal oversight than tactical decisions. The second thing is that this emphasis on the legal advisor is tenuous and somewhat ideological. IHL proponents must not be naïve about it, nor become complacent or satisfied either that this doctrine will always translate into practical application, or that it will necessarily survive a significant challenge from changing technologies that threaten this positive development in military doctrine.

3.3. The manual process of military decision-making

The MDMP is an analytical process that U.S. military commanders use to plan operations.⁵⁴ The methodology allows for the integration of many moving parts. It enables the commander, the planning staff and any other partners to properly understand the mission and the situation in which it is going to be carried out; to develop and compare various Courses of Action (COAs), decide on the best one and produce an operational plan and orders for subordinates to carry out.

According to the Operational Handbook for Judge Advocates practicing operational law in the U.S.,⁵⁵ JAs are required to be involved in every aspect of MDMP, including the plan development, and not just to review the plan at the end.⁵⁶ Although broken down into concrete defined steps, the actual process of MDMP can be complex, calling for judgment, experience and at times requires difficult choices. The JA needs to be privy to these in order to advise the planning staff of the legal implications of those choices during the process, to assist in developing a legally feasible plan⁵⁷ from the beginning. MDMP is often conducted under pressured and stressful conditions where time is of the essence and the Handbook states that “*increasingly, JAs serve as the “honest broker” in the review of plans and orders*”.⁵⁸ Addressing legal issues throughout the process thus saves valuable time and means that when it comes to reviewing the final plans and orders, the JA is thoroughly acquainted with the potential operational issues that may raise legal concerns.⁵⁹ So the more integrated the JA is in the process, the better.

A description of the current MDMP is below, including the job of the JA attached to the planning staff, in order to get an understanding of how much of the process is replaceable, practically and legally, by intelligent systems. The detail of each process, and the weight given to legal considerations, will depend on factors such as the level at which the MDMP is conducted, the circumstances and environment of the mission, and the personalities of the commander and planning staff. Whether a decision is a matter of strategy or tactics depends on the nature of the question, but the MDMP applies in principle to all forms of military decision-making. It also applies whether the planning is deliberate and in advance, with more lenient timelines, or under conditions of crisis and action-planning, where there is urgency, accelerated timelines and is demanding of a quick response, decisive action and flexibility to adapt on the ground. Although in the latter case the commander may order a truncated MDMP, it must be necessary and IHL continues to apply, even in a crisis.

Step 1: Receipt of the mission or task

The staff is told of the nature of the upcoming mission from a superior in the chain of command (such as headquarters) and begins to prepare to analyse it by gathering the necessary tools including plans from higher headquarters, maps of the area, field manuals and so on. The JA must ensure that all applicable legal materials are included in this initial preparation. Critical at this stage is to make an estimation of the time periods involved to allow sufficient time for subordinate commanders and

⁵⁴ Most modern organised militaries follow a comparable procedure

⁵⁵ Kouba, M., (ed.), *Operational Law Handbook*, 17th ed., (2017) The Judge Advocate General's Legal Centre and School

⁵⁶ *ibid* p. 495

⁵⁷ *ibid* p. 495

⁵⁸ *ibid* p. 499

⁵⁹ Such issues may extend beyond IHL to include other rules such as the states' Rules of Engagement, criminal code, and so on

staff to prepare for their operations. Depending on how time-constrained the particular operation is, the commander may elect to use a truncated MDMP.

Step 2: Mission Analysis

Here the commander begins a series of steps to establish 'battlefield visualisation' which involves (i) understanding the relationship of friendly forces to enemy forces and the environment; (ii) establishing the desired outcome for the whole operation (the commander's intent), and then (iii) determining the key tasks that will be necessary to achieve it. Each of these steps may give rise to significant legal concerns and it is the role of the JA to ensure that they are worked into the plan as they arise. To be able to do this effectively, the JA needs to have intimate knowledge of the parameters of the operation and the overall concept of the battle.⁶⁰

Step 3: Course of Action Development

The staff now needs to orient the mission by developing COAs based on the commander's guidance and intent for the mission. Typically, the staff will develop between two and five COAs, each one differing in, *inter alia*, the composition of forces, how various efforts are designated, and the use of resources and terrain. How many COAs are ultimately developed will depend on the commander's preference and time constraints.

A COA is a high-level specification of the operation represented by way of a COA sketch and COA statement, which together comprises a set of overarching actions, goals and sequencing of movements and objectives of friendly forces. The COA statement describes the concept of operations, which explains how the commander's intent will be accomplished by the various forces (including each warfighting function and information operations). The accompany COA sketch provides a visual description of the numerous aspects related to the specific manoeuvre. This step involves gathering together all possible intelligence and combining it with logistics to establish the most efficient and effect possible sequence. The JA must be a part of this process of developing the COAs so as to be fully informed about the legal implications of each one, and be in a position to advise the commander accordingly. For example, one COA may carry the forces through a densely populated area and lead to heavy indirect fire; another might require the damage of a dam to block the pathway for a possible counterattack; another may involve the use of Family of Scatterable Mines.⁶¹ Each of these scenarios has implications for IHL, involving a proportionality assessment, the destruction of civilian objects, and the choice of weapons respectively.

Throughout the development of the COA, the staff will generally use a synchronisation matrix to record the content of the plan. As its name suggests, it is a way of coordinating the various tasks and actions of the multiple subunits and assets of the friendly force into different blocks which are viewable against the expected time periods for each action and can establish the precise moments for decisions are to be made *in casu* depending on the tactical situation and enemy actions and reactions. It will also include the objectives and manner of execution of each action, information about routes and locations, availability of supplies, estimated combat losses and so on. The synchronisation matrix is a convenient way to display the information in order to identify the range of actions and points of coordination involved in the operation, and to be able to conduct the subsequent step: namely to ascertain the feasibility of a COA and its likelihood of success.

Step 4: COA Analysis

This step is what is known as wargaming, which is a technique for visualising the flow of a battle to determine the feasibility of each COA. The staff tests each COA against potential enemy actions and reactions to produce an analytical baseline that the commander ultimately uses to decide on the

⁶⁰ Kouba (n. 55) p. 497

⁶¹ Kouba (n. 55) p. 498

best COA.⁶² By simulating the COAs, it allows for each one to be examined in detail and for a comprehensive plan to be developed, while at the same time recording the strengths and weaknesses of each respective COA. Experiments with wargaming estimate that the process generally takes a team of people approximately two to eight hours for each COA.⁶³

The strengths and weaknesses of a COA would concern practical advantages and disadvantages – such as efficient scheduling and allocation of tasks, assigning suitable locations and the outcome of attrition estimates – but it should also consider the legal issues that arise in respect of both friendly and enemy forces' actions. Thus the JA plays a vital role during the wargaming process in that it is the ideal opportunity to address such issues as and when they arise so that they can be built into the overall assessment of the COA. For example, a particular COA may encourage the particular enemy force to react by using poisonous gas in a counterattack. With the help of the JA, the staff could then build in its reaction to this beforehand to ensure its response is not, in turn, a violation of IHL. The JA would be able to advise whether a reprisal is legally permissible in the particular circumstance. The analysis requires creativity, knowledge and experience. It is also subject to the will of the commander and all the particular biases and priorities of the planning staff, commander and country or armed force.

Step 5-6: COA comparison and Approval

Having determined the advantages and disadvantages, the staff then evaluates and rates the various elements of each COA using a defined set of criteria that has been established prior to the wargaming process, compares the results and selects the best COA. A synchronised operational plan is then composed out of the preferred COA, together with the staff's recommendations regarding the preferred COA. The commander selects the best COA, proposes amendments or rejects them all and orders the staff to begin the process again. How 'best' is defined in these situations is a matter of context. It may mean accomplishing the commander's intent with the greatest chance of success, the most effective use of resources,⁶⁴ or the most compliant with the rules and principles of IHL. Ideally, it will be a balance between all of these considerations, but how they are weighted will depend largely on the commander, and the tactical situation on the ground.

The JA's involvement in the earliest part of this process is therefore crucial in the determination of whether a COA is legally supportable. The Handbook emphasises the importance of the JA identifying any legal issues before the staff expends valuable time and resources in planning it.⁶⁵ It is clear that the less involvement the JA has during the earlier phases of MDMP, the greater the loss of any tactical advantage resulting from the effective use of resources. Another consequence, which is not expressed in any handbook, is that input from the legal advisor at the later stages is probably more likely to be ignored because of the time pressure and priorities of the commander.

Step 7: Orders Production

The final COA is refined and turned into a plan, which is a precise statement of the concept of the operation and an explanation of all the information needed to support it. The JA should review the final plan and the resultant orders to subordinates, something that he or she will be able to do much better and more efficiently if there has been proper involvement at each step of the MDMP.

⁶² U.S. Army Field Manual 101-5, *Staff Organization and Operations*, (1997)

⁶³ Rasch, R., Kott, A., and Forbus K., *Incorporating AI into Military Decision Making: An Experiment* (2003) IEEE Computer Society

⁶⁴ Ground (n. 6) p. 5

⁶⁵ Kouba (n. 55) p. 499

3.4. Superior Responsibility: The interface between IHL and warfare

There are two avenues by which criminal sanctions are imposed on superiors (and other individuals) for war crimes and grave breaches of IHL: first, by direct responsibility for intentionally planning, ordering, committing, aiding and abetting or participating in a joint criminal enterprise, and second, by indirect responsibility for failing to prevent and punish the commission of war crimes and grave breaches of IHL by subordinates under their effective control. There are also additional sanctions, not necessarily criminal, for violations that do not meet the standard of grave breaches or war crimes, which are almost exclusively applied through domestic courts and military disciplinary processes.

Superior responsibility is grounded in the principle of responsible command, which is a general principle of IHL that both underlies the doctrine of superior responsibility, and serves as an important interpretive tool by which the scope of the doctrine is understood. The ICRC explains it thus

“the role of commanders is decisive [...] the necessary measures for the proper application of the [Geneva] Conventions and the [Additional] Protocol must be taken at the level of the troops, so that a fatal gap between the undertakings entered into by Parties to the conflict and the conduct of individuals is avoided. At this level everything depends on commanders, and without their conscientious supervision, general legal requirements are unlikely to be effective.”⁶⁶

The doctrine of superior responsibility, grew out of a recognition, initially by international courts and tribunals and later codified in a number of treaties, that superiors are uniquely positioned at the intersection of war and law and, because of the control they exercise over their subordinates, the effective implementation of IHL in armed conflict depends considerably on commanders ensuring that they and their subordinates respect it at all times.⁶⁷

The principle of responsible command is, therefore, a general concept regarding the duties of superiors to adopt measures that secure general adherence to IHL through well-functioning military, political and other structures.⁶⁸ The law of superior responsibility, on the other hand, is derived from this principle but is concerned primarily with holding military superiors accountable for breaches of specific, legally binding duties that carry criminal sanctions.⁶⁹ Beginning with the trial of General Tomoyuki Yamashita by a U.S. Military Commission in Manila following WWI, the doctrine began as a mechanism for holding high-ranking superiors criminally responsible for the conduct of their subordinates even where they were neither involved in nor aware of the crimes. Despite an inauspicious beginning, it was subsequently refined and forged into a workable doctrine by the various ad hoc international tribunals, particularly in the jurisprudence of the International Criminal Tribunals for the Former Yugoslavia and for Rwanda and has now been codified by international treaties. Article 87 of Additional Protocol I explicitly ascribes a duty to military superiors to “*prevent and, where necessary, to suppress and to report to competent authorities breaches of the Conventions and of this Protocol*” and The ICC Statute assigns superior responsibility to military commanders or those acting as military commanders with effective control over subordinates that have committed crimes as a result of his or her failure to exercise that control properly. In short, a superior will be criminally liable in those circumstances if (i) the superior knew or should have

⁶⁶ ICRC, *Commentary on the Additional Protocols*, (1987) para. 3550

⁶⁷ *ibid*; Mettraux (n. 11) p. 5

⁶⁸ Mettraux (n. 11) p. 55

⁶⁹ See for example, *Prosecutor v Hadžihasanović et al*, Case No. IT-01-47-A, Judgment 22 April 2008, Article 7(3) AC Decision, para 22 in which the Appeals Chamber recognised that “there is a difference between the concepts of responsible command and command responsibility. The difference is due to the fact that the concept of responsible command looks to the duties comprised in the idea of command, whereas that of command responsibility looks at liability flowing from breach of those duties. But as the foregoing shows, the elements of command responsibility are derived from the elements of responsible command.”

known about the planning or commission of a serious violations of IHL, and (ii) failed to prevent such crimes or to punish those that committed them.⁷⁰

An individual subject to superior responsibility means anyone that is in effective control over the conduct of subordinates in an organised military structure, not necessarily a state military. Such individuals may be engaged in high-level strategy, in operations, and even non-commissioned officers that have effective control over only a few people.⁷¹ This broad definition means that the body of law can adapt to encompass future scenarios as the character of armed conflict changes, as traditional lines between roles are blurred, and technology gives rise to unpredictable developments.

Consequently, the basis of superior liability lies in the failure of the superior to prevent crimes being committed by subordinates or, put another way, to ensure that subordinates respect the norms of IHL during the conduct of warfare. Implicit in this liability lies an extensive set of duties applicable to superiors throughout the command hierarchy to take preventative measures against violations of IHL. The doctrine, therefore, has two primary objectives; first, to institute a set of standards that support the proper functioning of any chain of command and that guarantee respect for IHL norms within that chain of command, and second, to establish a form of criminal liability and a measure by which to prevent specific crimes being committed by subordinates within the chain of command.⁷² Although subject to significant debate about the scope of the doctrine, on the whole modern superior responsibility is considered an important mechanism for addressing the culture of impunity that often accompanies war involving widescale atrocities.⁷³ This wider understanding of the doctrine of superior responsibility, which sees it as going beyond merely a legal basis for the prosecution of superiors who fail to prevent war crimes, towards *an internationally sanctioned minimum standard of conduct* that requires superiors to take responsibility for controlling the prosecution of war and ensuring it is done within the limits of IHL,⁷⁴ is vital for understanding how it is implemented.

The implementation of superior responsibility is best understood in light of IHL more broadly, which is fulfilled by way of a three-pronged approach of prevention, control, and suppression. The latter category constitutes *ex post facto* measures such as state and individual sanctions and depends largely on the political will of states to enforce it, particularly in respect of the prosecution of individuals for war crimes and serious violations. Opponents of IHL often concentrate on the weakness of its accountability mechanisms to criticise the effectiveness of the entire body of law, and even its stoutest proponents are aware of the difficulties involved in holding individuals criminally liable for their conduct during war.⁷⁵ Consequently, preventative mechanisms take on an additional significance and much of the ability of IHL to achieve its objectives depends on the consensus of states and other actors to implement its provisions of their own accord, by establishing duties on its actors, incorporation into domestic rules, military manuals, and rules of engagement, and through the use of legal advisors.⁷⁶ The same can be said of the responsibility of superiors, and so the ICRC Handbook on command responsibility and U.S. military doctrine have laid out the positive duties on persons exercising command.

⁷⁰ ICC Statute, Article 87 available at <https://www.icc-cpi.int/resource-library/Documents/RS-Eng.pdf>

⁷¹ ICRC, *Commentary on the Additional Protocols*, (1987) paras. 3553-4. This broad definition is intended to include non-military commanders, military commanders, and individuals acting like military commanders

⁷² Mettraux (n. 11) p. 18

⁷³ Mettraux (n. 11) p. 16

⁷⁴ Mettraux (n. 11) p. 20

⁷⁵ Kolb (n. 8) p. 187

⁷⁶ Kolb (n. 8) pp. 189-190

The framework of IHL rules and principles to be taken into account during the decision-making process

Part of the reason that regulating the use of AI in decision-making and planning processes will be complicated is that there are no treaty or customary rules that explicitly governs how commanders conduct their activities, only that they are required to ensure that the overall mission and general strategy of the armed conflict is carried out in a manner that is consistent with IHL.⁷⁷ There is thus no clear legal basis on which to ground a general prohibition on the use of this technology in military decision-making.

According to the ICRC, the obligations of commanders are broken up into stages that chart similar lines as the implementation of IHL through prevention, control and suppression. There are three loosely established categories of duties.⁷⁸ First, those duties relating to ensuring the troops are aware of the content of their own obligations to respect IHL, commensurate with their function and the specific mission, through training and consolidation into the structures, doctrine and culture of the armed forces, which takes place primarily during peacetime. Second, the specific measures that a superior is required to take on becoming aware that a violation is about to occur, or has already occurred, within his or her sphere of control, both during and post armed conflict. Third are the general measures that are required to prevent violations from occurring during an armed conflict through, *inter alia*, ensuring that the commanders own orders are capable of being carried out in a manner that is consistent with IHL.⁷⁹ This latter category is crucial in the planning and decision-making process during armed conflict.

First Leg: Preventative measures related to the planning and decision-making tasks.

The planning and execution of warfare converge largely at the operational level, where the high-level strategic objectives of the state are linked with the tactical objectives on the ground. Concerning the substantive provisions of IHL, in principle, the planning and decision-making process encompasses the entire body of IHL norms, including the general principles, and the myriad of rules that are derived from them and expressed in treaties, customary IHL, court decisions and domestic laws and rules of engagement. Much will depend on the precise nature and size of the operation to determine which principles or laws apply to a particular engagement, including whether and to what degree the issue in question relates to the means and methods of warfare, and which category of persons and objects stand to be affected. It is crucial therefore that commanders at each level know the law relevant to his or her responsibility and, if necessary, has advisors to that end. Given the vast and complex system of such rules, the importance of military legal advisors in the planning process is apparent, and their role in modern armed forces significant, although their level of engagement will depend on the organisation and ideology of the armed forces in question.

Commanders are responsible first and foremost for ensuring that his or her decisions and orders comply with the provisions of IHL.⁸⁰ Specific rules will be triggered at different stages of the planning cycle and will depend to some extent on the availability of time,⁸¹ which can be decisive in obtaining compliance with IHL. During pre-planned operations, in which there is time available and which are most likely to take place at higher levels of strategic and operational decision-making, the law requires “*every care to be taken to include the laws of armed conflict into the planning cycle.*”⁸² During improvised actions on the other hand, which are subject to severe time constraints,

⁷⁷ Additional Protocol I, Articles 1, 80, 87

⁷⁸ ICRC, *Commentary on the Additional Protocols*, (1987) paras. 3552-3556

⁷⁹ ICRC, *Handbook on International Rules Governing Military Operations*, (2013) p. 259

⁸⁰ *ibid*; Mettraux (n. 11) p. 270-271

⁸¹ *ibid*

⁸² Additional Protocol I, Article 57; ICRC Customary IHL Rules 15-21

immediate tactical concerns take priority. Nevertheless, commanders are still required to respect IHL at all times, and to take all measures required by IHL which the tactical situation permits.⁸³

There are in fact two scenarios where the law explicitly permits commanders to make legally important decisions based on a subjective assessment of the situation on the ground wherein they are required to evaluate whether a proposed action meets the requirements in respect of each situation. The first is where a particular provision of IHL makes a commander's duty to act or to refrain from acting in a certain way dependent on the tactical situation on the ground, phrased either as "*if the tactical situation permits*" or as "*if the tactical situation makes it impossible*". The tactical situation, therefore, becomes a factor in the commander's decision whether and how far to apply that rule.⁸⁴ The tactical situation, however, comprises all circumstances, including military and humanitarian considerations. The other situation refers to those IHL rules which are generally applicable but permit for derogation or waiver where there is a degree of military necessity aimed at achieving a legitimate military purpose. However, because the principle of military necessity is baked into the very notion of IHL and each rule and principle, derogation from a rule on the basis of military necessity is necessarily confined to a limited set of circumstance. For example, the general prohibition against the use of cultural property for purposes that are likely to expose it to destruction or damage may be waived in the face of imperative military necessity. Another example is the prohibition against destruction of objects that are indispensable for the survival of the civilian population, which may be derogated from where it is the forces own territory, and only as a defensive measure where there is an imperative military necessity.⁸⁵

Both of these scenarios involve difficult, subjective assessments and the exercise of the independent judgment of the commander who must evaluate whether the particular situation meets the required threshold for deviation from the law, and whether to take a particular action or not. The ICRC Handbook calls this the 'estimate' or 'appreciation process' (essentially the MDMP in U.S. military doctrine) which concerns operational planning either during or in anticipation of armed conflict. Based on all available information and his or her appreciation of the facts, the commander prepares an 'estimate' that is to be the basis for determining all potential options and alternative actions. The estimation process itself must allow for consideration of IHL to be taken into account, and for estimates to be modified accordingly.

It has been noted that the MDMP and the orders stemming from it must take into account all applicable law, including those applicable to any specific contextual obligations, such as laws relating to occupation, United Nations peacekeeping operations, or the laws of neutrality. While not practical to present every rule, below is a brief outline of the most crucial rules and principles that apply to the MDMP in armed conflict at any level but especially at the operational level.

To begin with there is a general duty on superiors to make good use of intelligence to ensure clear, accurate and up-to-date knowledge of the tactical situation, including the number and density of civilian groups and civilian objects. Then there are the obligations that arise out of the following principles of IHL that operate both as general principles and operational rules.

- Humanity

The principle of humanity enshrined in the Martens clause is a broad principle intended to encompass all situations of armed conflict and any involvement therein. Any decision that involves the use of force and loss of life requires a qualitative judgment and depends on compassion, intuition and an ability to evaluate particular context. It is contained in Article 1(2) of Additional Protocol I,

⁸³ Additional Protocol I, Article 57; ICRC Customary IHL Rule 19

⁸⁴ Such as providing an effective warning (unless tactical situation does not permit); movement of different categories of transport through separate routes (unless the tactical situation does not permit), search for and collection of victims (as soon as the tactical situation permits)

⁸⁵ Additional Protocol I, Article 54(2); ICRC Customary IHL Rule 54

the preamble to 1907 Hague Convention (IV) and was confirmed by the International Court of Justice to be “*an effective means of addressing rapid evolution of military technology*”⁸⁶ and representing customary international law.⁸⁷

- Limitation⁸⁸

The principle of limitation is found in the text of the Hague Regulations, Article 22, which says that “[t]he right of belligerents to adopt means of injuring the enemy is not unlimited” and again in Additional Protocol I, Article 35(1) which holds that “[t]he right of the Parties to the conflict to choose methods and means of warfare is not unlimited”. It is the principle that negates total war, serves as a brake on military permissiveness, and in fact is the essence of IHL, underlying every restrictive rule of IHL, especially in regards to means and methods of warfare.

- Distinction

The principle of distinction holds that an attack which does not properly distinguish between protected persons and objects⁸⁹ and legitimate military targets is indiscriminate and unlawful. It serves to limit the effects of war by channelling the use of force against legitimate military targets, and by providing civilian persons and objects with immunity from attack. The principle of distinction, in turn, gives rise to specific obligations on actors taking part in the conflict. For example, Article 38 of the Fourth Geneva Convention establishes the Red Cross, Red Crescent, Red Lion with Sun and the Red Crystal as protective symbols to be used during armed conflict to mark all persons and objects protected by the four Geneva Conventions and it thus falls to commanders to factor all protected persons and objects into their decision-making and to recognise the use of these emblems.

- Precaution

As a consequence of the principle of distinction, superiors are required to take precautionary measures to avoid targeting protected persons and objects during their preparation for conflict and planning of operations, something of such importance that it warrants being a self-standing principle.⁹⁰ The degree of precaution required under IHL will depend on the specific circumstances, but at a minimum it requires that an attack must be carefully prepared with an aim to sparing civilian life and objects as much as is feasible.⁹¹ ‘Feasible’ here means everything that is practically possible, taking into account all the circumstances at the time, including military and humanitarian considerations. The choice of means and methods must aim at ensuring that an attack does not cause excessive collateral damage in relation to the concrete and direct military advantage anticipated.⁹² It is also essential that precautionary measures be communicated efficiently down the chain of command.

- Proportionality

The notion of sparing civilian life and object ‘*as much as is feasible*’ lies at the heart of the principle of proportionality. It is the principle which best reflects the compromise of IHL as it requires balancing the damaging consequences of an action with the importance of the military objective. An attack that causes excessive collateral damage in relation to its military objective is thus expressly

⁸⁶ *Case concerning the Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion*, ICJ Reports, 1996, para 78

⁸⁷ *ibid* 84

⁸⁸ Kolb (n. 8) p. 80

⁸⁹ Additional Protocol I, Article 48; and ICRC Customary IHL Rules 1 and 7 respectively

⁹⁰ Additional Protocol I, Article 57; ICRC Customary IHL Rules 15-21

⁹¹ Kolb (n. 8) p. 81

⁹² Additional Protocol I, Articles 53 and 58

prohibited under IHL.⁹³ There are three facets to this rule: (i) the action must fulfill the aim sought, (ii) of several potential actions that would achieve similar results, the least onerous must be selected, and (iii) the action taken must be commensurate with the military objective,⁹⁴ which assessment is quite clearly easier said than done and, in the fog of battle, the ability to conduct this assessment properly is one of the defining features of a good commander. The assessment is difficult in that, again, what constitutes ‘excessive’ and ‘commensurate with the military advantage’ are highly subjective determinations. This rule gives rise to positive obligations on commanders under IHL to, *inter alia*, take precautionary measures in respect of the choice of weapons, coordination and timing of operations, and provision of advanced warnings.

- Military Necessity

Depending on the ideological viewpoint of the source, the definition of military necessity differs. The ‘orthodox’ international law perspective sees it as a dual principle, both a permissive and constraining, which sanctions the use of force insofar as it is necessary to achieve a military objective, and no more. At the operational level of warfare, this means only using violence that is useful and the least destructive option available.⁹⁵ However as noted above, military necessity is written into the inherent structure of the whole body of IHL and thus cannot also function as a justification for failing to comply with those laws, except in the limited circumstances whereby a rule expressly allows for this analysis by an individual in the battlefield. These rules are subject to a hierarchy which is useful for the purposes of interpretation.⁹⁶ Under the Geneva laws, there is no exemption allowed for military necessity, but there may be an exemption for *imperative* military necessity. For example, the obligation of a protecting power to supply food and medical supplies to the population may be temporarily suspended where imperative military necessity requires.⁹⁷ The Hague Laws, on the other hand, allow for some exemption for military necessity or need, for example concerning the facilities that must be accorded to relief societies for prisoners of war.⁹⁸

The U.S. Army Field Manual provides that “*Military necessity has been defined as that principle which justifies those measures not forbidden by international law which are indispensable for securing the complete submission of the enemy as soon as possible.*” Besides revealing an ideological tendency towards viewing IHL as a predominantly permissive system, this seemingly straightforward definition is complex in the context of warfare. ‘Indispensable’, ‘complete submission’, and ‘as soon as possible’ are all subjective terms that require a degree of careful assessment at critical moments.⁹⁹

In addition to complying with the above principles, operational superiors are required to ensure that orders to their subordinates (which ultimately flow from the MDMP) are practicable and can be carried out in conformity with the law, and must provide precise instructions in respect of (i) where protected persons and objects are located, (ii) rapid evacuation of wounded and treatment of POWs and respecting and collection of dead, including the enemy, (iii) under which circumstances to open fire and which weapons to use, (iv) subject to the element of surprise, cooperation with civilian authorities in order to reduce the risk to civilian persons and objects,¹⁰⁰ such as warning about specific routes and shelters, providing evacuation warnings and if necessary assistance, and liaising with medical authorities,¹⁰¹ (v) respecting and protecting civil defence systems, and (vi) allowing for the passage of humanitarian relief.

⁹³ Additional Protocol I, Article 51(b); ICRC Customary IHL Rule 14

⁹⁴ Kolb (n. 8) p. 76

⁹⁵ Additional Protocol I, Article 57 (3)

⁹⁶ Kolb (n. 8) pp. 87-88

⁹⁷ Geneva Convention IV, Article 55(3)

⁹⁸ Hague, IV, Article 15

⁹⁹ Kolb (n. 8) pp. 83- 92

¹⁰⁰ Additional Protocol I, Article 58; ICRC Customary IHL Rules 22-24

¹⁰¹ Additional Protocol I, Articles 57 and 58

Lastly, commanders must maintain control over the execution of their orders, ensuring that there is continued respect for the law, to interrupt action as and when it appears that a violation of IHL will occur, and finally to hold accountable those who have violated it. The next question that arises then is how does IHL hold those commanders accountable for the failure to meet their obligations?

Second leg: Accountability

- Indirect superior responsibility

In addition to the abovementioned duties, superior responsibility is also a mechanism of indirect liability, by which superiors, military and non-military, can be held criminally liable for failing to prevent or punish the commission of war crimes by subordinates under his or her effective control or effective authority.¹⁰² It is governed by customary IHL and under the ICC Statute, the upshot being that there is a doctrinal dispute about the precise nature and scope of the responsibility of commanders in these circumstances.¹⁰³ Under the ICC Statute, the basis for superior responsibility is the “failure to exercise control properly” over those forces that have committed unlawful acts. It is therefore based on liability for an omission in that it arises from dereliction of duty rather than strict liability for the crime committed. While providing clarity on this point, the ICC Statute creates confusion on the question of *mens rea* by establishing two separate regimes for non-military and military superiors,¹⁰⁴ with two different standards of liability. Military superiors are subject to a lower threshold in that they are responsible for crimes of which they “knew or, owing to the circumstances at the time, should have known.” Non-military superiors, on the other hand, are only responsible if they “knew or consciously disregarded information which clearly indicated” that a crime was about to or had been committed. This higher threshold matches the customary law,¹⁰⁵ which it applies to both military and non-military superiors, regardless of their level or function.¹⁰⁶ The positive duty of superiors to take necessary and reasonable measures is divided into three obligations namely to prevent, repress, and submit crimes to a competent authority to investigate and prosecute.¹⁰⁷

- Direct criminal responsibility

Lastly, individuals can be held directly criminally responsible for intentionally planning, instigating, ordering, committing, aiding and abetting or participating in a joint criminal enterprise.¹⁰⁸ International criminal law establishes the contribution of individuals to war crimes or other serious violations, based on a recognition of the fact that they are typically collective in nature in that they involve a large number of persons cooperating through a more or less organised network. Furthermore, the degree of responsibility does not necessarily diminish the further away an individual may be from the commission of the particular crime. If anything it increases, particularly if the persons in question hold higher positions in the chain of command.¹⁰⁹ Individual responsibility for a crime can thus be established either by the commission of the crime (as an individual, part of a joint commission or commission through another person)¹¹⁰ or by instigation or ordering the

¹⁰² ICC Statute, Article 28(a)

¹⁰³ In sum, the dispute concerns whether or not superior responsibility constitutes a *sui generis* mode of liability for omission linked to the individual’s status as a superior.

¹⁰⁴ Article 28(a) and 28(b) respectively

¹⁰⁵ Although it is usually phrased slightly differently as “knew or had reason to know”. Although the two align in this respect, the position under the ICC Statute is not considered to reflect the customary IHL as a whole

¹⁰⁶ Mettraux (n. 11) p. 31

¹⁰⁷ ICC Statute, Article 28(a)(ii) and 28(b)(iii). Similarly, Addition Protocol I, Article 87 and, more generally, the case law of the Ad Hoc International Criminal Tribunals for clarification on the parameters of the duty to prevent and punish the crimes of subordinates

¹⁰⁸ ICC Statute, Article 25(3), *also see* ICTY Statute, Article 7(1); ICTR Statute, Article 6(1)

¹⁰⁹ Werle, G. and Jessberger, F., *Principles of International Criminal Law*, (2014), p. 193

¹¹⁰ ICC Statute, Article 25(3)(a)

commission of a crime.¹¹¹ In respect of liability for ordering the commission of a crime, this assumes the existence of a superior-subordinate relationship whereby the perpetrator uses his or her position of authority to induce the subordinate to commit a crime. The order need not be explicit but the persons giving it must intend that a crime will be committed on execution of the order, or at least be aware that there is a substantial likelihood that a crime will be committed,¹¹² even if the order involves an act that is not *per se* prohibited under international law.¹¹³ Under the ICC Statute, the standard for *mens rea* is even lower. In that case it is sufficient for the person giving the order to be aware that the subordinates' intention to commit a crime, without necessarily sharing in it. It is a reflection of the hierarchical structure of authority, and a recognition that those higher up the chain of command bear a greater responsibility for being an abettor of to a crime, although the crime does not constitute a form of principal liability. It is intended to capture those commanders at the higher and intermediate level of command that receive and issue orders.¹¹⁴

3.5. Conclusion

Military superiors are the guardians of the application of IHL rules in the field of battle. The ICRC Handbook instructs them to incorporate the rules of humanity throughout the whole conduct of war: from the inception of the concept of the war, through the numerous aspects of planning and execution. Integration of the law into the operational practice in the field is the key to ensuring that there is compliance throughout the entire chain of command from strategy, through the operational to the tactical level.¹¹⁵

Criminal liability of superiors is limited to the commission of war crimes and grave breaches of IHL and, while there are fairly extensive definitions of these categories, much of the conduct of commanders in MDMP will fall short of being crimes, and thus require other means of implementation. The emphasis placed on the legal advisors in the MDMP indicates the significance of this implementation through preventative measures.

The picture that emerges from this vast set of responsibilities and considerations, not to mention the many military considerations at play, is one of overall awareness and control. Besides giving rise to the obligations discussed here in respect of operational decision-making, it is this notion of awareness and control that lies at the heart of the indirect criminal responsibility mechanism of superior responsibility. In the case of direct criminal responsibility, particularly of high-ranking military and non-military superiors, the basis is something else. Rather than trying to connect them directly to the commission of a particular crime, which in reality is extremely difficult to do, it is based on a more fundamental denial of impunity, and a desire to hold those leaders accountable who get others to implement their evil agendas. The difficulties associated with accountability mechanisms in IHL, however, means that additional emphasis is placed on prevention mechanisms than would be the case in a domestic criminal justice legal system.

¹¹¹ ICC Statute, Article 25(3)(b); *also see* ICTY Statute, Article 7(1); ICTR Statute, Article 6(1); Geneva Convention III-IV, Articles 129(1)/ 146(1)

¹¹² Werle and Jessberger, (n. 109) p. 215

¹¹³ This is according to the ICTY in its judgment of *Prosecutor v Boškoski and Tarčulovski*, Case No. IT-04-82-A19, Judgment 19 May 2010, AC Decision, paras 172-174 but has been criticised by some international law scholars

¹¹⁴ Werle, G. and Jessberger, F., *Principles of International Criminal Law*, (2014), p. 216

¹¹⁵ See generally Chapter 8, 'Command Responsibility', in ICRC, *Handbook on International Rules Governing Military Operations*, (2013) p. 259; Mettraux (n. 11)

SECTION IV - FUTURE PARADIGM: AI AND MDMP

4.1. CADET - A test case of a modular AI planning system

The Course of Action Development and Evaluation Tool (CADET)¹¹⁶ is presented here as an example of how even relatively basic AI systems can potentially affect the implementation of IHL during the military decision-making process. An understanding of the capabilities of current and near future AI (presented herein following this case study) reveals that this impact will be increasingly significant, in ways both positive and negative.

As its name suggests, CADET is a computerised tool that assists military planning staff to develop and assess the best possible COA for a given mission. Briefly, it employs AI techniques to translate a high-level battle concept, or mission, into an actionable plan in a two a step process. First it translates the high-level concept into various COAs, essentially by breaking it down into hundreds of detailed tasks that are necessary to carry out the mission and achieve the intended objective. Then it determines the feasibility of each COA by testing it against the most likely and/or the most dangerous enemy COAs i.e. it conducts the wargaming analysis. Computationally, this is achieved by an overarching algorithm for interleaving incremental planning, adversary estimates, scheduling, routing, estimating consumption and attrition, and so on), which essentially means that each element is added in a manner that intersects and overlaps with the other. The interleaving technique essentially allows for several algorithms to be used in combination by switching and alternating between them. It emanates from a knowledge-based approach, this case applied to adversarial reasoning, whereby the system plans backwards from an overarching set of instructions.¹¹⁷

Although CADET was tested in an experiment specifically at the brigade level, it is notable in that the technology it employs is generic and thus can be applied to a range of different functions – particularly those that require adversarial reasoning and detailed planning in constrained time periods. The specific technical approach is applicable to any explicit adversarial reasoning exercise whereby potential adverse effects of actions and the appropriate precautions are planned.¹¹⁸ It is a knowledge-based tool meaning that it can be adapted to new domains, all that need happen is that the knowledge base is changed. Although the test described here regards the brigade-level MDMP, which is the lowest level of headquarters that conducts this kind of planning, the generic nature of the technology means that in principle it could be used higher up the hierarchy, and across functions and domains. The same authors of CADET have also built demonstrations whereby the tool is used for other tasks including intelligence collection using scouts and unmanned aerial vehicles; combat tasks of robotic forces, and responses to terrorism incidence in urban environment.¹¹⁹ More recently they have demonstrated its use in coalition actions.¹²⁰

The system is designed to resemble the manual process of MDMP as closely as possible. One of the human planning staff defines the high-level concept of the operation via a human-machine interface, by way of a COA sketch and a COA statement. The COA interface transforms this collection of formal assertions and/or objectives for the mission into an electronic representation that is the input for the CADET system proper, which then breaks in down into the relevant various attributes that together form a high-level scheme of maneuver for the unit to carry out. Examples of such attributes, although there are many others, would be (i) the number and character of friendly and enemy units;

¹¹⁶ Kott, A., Budd, R., Ground, L., *et al.*, *Decision Aids for Adversarial Planning in Military Operations: Algorithms, Tools and Turing-test-like Experimental Validation*, (2016), CULAIJ

¹¹⁷ *ibid*, p 6

¹¹⁸ *ibid*, p. 29

¹¹⁹ *ibid*, pp 7-8

¹²⁰ Ground, L., Kott., A., Budd, R., *Coalition-based Planning of Military Operations: Adversarial Reasoning Algorithms in an Integrated Decision Aid*, (2015), CoRR

(ii) any control measures such as area boundaries; (iii) the network of passable terrain to be covered; and (iv) a set of activities - including estimates of enemy activities - and the temporal relationships between them. Each attribute is ascribed a certain value in order to determine its priority in the planning system. This particular technology allows for it either to attach default values in respect of each attribute, or there is a place for a human planner to explicitly specify values.

CADET then translates this high-level 'definition' of the COA into a detailed battle plan. It breaks down tasks into detailed actions; determines and allocates additional support tasks to friendly assets; estimates enemy forces possible actions and reactions and devises counter-actions for friendly forces; and estimates elements such as timing, supply consumption and attrition. This automated process takes approximately 20 seconds, as opposed to the several person-hours that a team of human planners usually takes. It results in a significantly detailed, wargamed plan for the unit that is represented in a synchronisation matrix (SM), which a human planner is then able to review and make amendments, either to the original high-level specification or directly to the plan. Finally, a satisfactory plan is reached, the whole process taking between five and thirty minutes, and the COA analysis can be presented to the commander who produces operational orders for the unit.¹²¹ The true value of CADET lies in its ability to conduct these calculations at speeds unattainable by human planners, as the experiment below reveals.

4.2. The Integrated Course of Action Critiquing and Elaboration System

Although conducted some years ago, the ICCES experiment¹²² is shown here as an example for two reasons. First, it is the kind of amalgamation of various AI techniques into one system, that works together with human planners, that will most likely be the near-future of military planning. Second, as will be discussed presently, the U.S. has not pursued research and development in this area over the last decade and so there have not been many applications or research to replace it.

CADET was tested in a collaborative experiment between Northwestern University, BBN Technologies¹²³ and the U.S. Army in 2003, which sought to determine how AI-based tools can be incorporated into MDMP. It brought together several prototype technologies based on AI techniques, including CADET, to create a set of decision aids which were then tested by Army officers in controlled experiments. The authors noted that there is a big dichotomy between the planning and execution of operations. Very often the critical decision-making that inevitably needs to take place during execution of a plan (due to unforeseen circumstances, or failures) does not resemble the original plan. The whole planning process is too slow and cumbersome for critical decisions. The experiment sought to integrate these two things better, more efficiently using automated systems to achieve fast, real-time decisions and re-planning.

The Experiment

The ICCES was modeled on the manual MDMP described in Section II, which commonly takes two to eight hours, involves considerable numbers of staff personnel and often occurs in difficult, stressful physical environments, but focused specifically on COA planning and analysis at the Army Brigade level. Planning here involves documenting and communicating high-level COA to producing a detailed analysis and plan of tasks. It did not include the creative step of inventing a high-level COA.

¹²¹ Kott, Budd and Ground (n. 116) pp. 3-6 for a comprehensive explanation, and Figures 1-6 for visual representation of the process

¹²² Rasch, R., Kott, A., Forbus, K., *Incorporating AI into Military Decision Making: An Experiment*, (2003) IEEE Computer Society, p 18

¹²³ BBN Technologies, Now Raytheon BBN Technologies, see <https://www.raytheon.com/>

Existing AI prototype tools (from advanced R&D), were modified and integrated loosely into a suite of decision aids that worked together (the 'Rig') which was then used to conduct experiments around the key tasks of the staff. The following tools were used:

COA Sketch Creator tool that enables the user to sketch a COA into the computer, which stores the semantic, knowledge-based representations for each item users add to the COA sketch. It uses AI statistical recognisers for visual symbols, speech and natural language understanding and then integrates these information sources into commands for the underlying software system.

COA Statement tool allows the staff to enter the COA statement by using templates and tightly constrained grammar to formulate sentences that can be parsed into internal representations. This tool automatically fills in some templates with elements derived from the sketch, such as units and tasks, so that the sentences now refer to the sketch.

Fusion Engine that merges the COA Creator and Statement tools to generate a single information file and eliminate inconsistencies between the information. It then uses its XML translator to translate the fragments of knowledge into the XML schema, which is a digital representation of the sketch and statement information.

CADET translates and transforms the XML schema into a detailed battle plan, including information regarding logistic and attrition calculations, as described above. This is capable of being done automatically or with human guidance.

The set up therefore resembled the manual process of entering a COA sketch and COA statement, reviewing the detailed plan, and modifying the sketches and statements accordingly.

- The Primary Experiment

Eight Army officers from combat arms branches, with tactical experience ranging from 11-23 years of active experience, were divided into two groups. After being trained on how to use the system they were given the same tactical scenario and high-level plan from 'headquarters' and required to conduct the MDMP, one manually, the other using ICCES. Each had to deliver a COA sketch, statement and synchronisation matrix reflecting the best COA with sufficient detail for the plan to be executed. The following day the groups were given a new scenario and their tasks swapped around.

- The Auxiliary Experiment

In the second experiment, five difference scenarios were taken from U.S. Army exercises which were all brigade sized, offensive actions, that differed in terms of terrain, mix of friendly forces, nature of opposing forces, and scheme of maneuver. The scenarios were fed into ICCES, and CADET automatically generated detailed plans, expressed as SMs. A human user made edits to the SM where were minor but nevertheless there was a human to supervise the final product. The whole process took approximately twenty minutes per SM. The estimate of manual production of each scenario was approximately sixteen person-hours were needed per planning product, so the time saving aspect was already significant.

Nine judges, mostly active duty officers of appropriate rank and experience, were then given a package containing the sketch, statement and SM plus attrition and consumption estimates for each scenario, , to assess and assign a grade between 0-10 (5 being comparable to current typical products). The results showed that the CADET won 2, lost 2 and tied 1. Essentially the resultant COAs received the same overall grade but in a substantially reduced time-frame, and with far few human planners involved.

- Tactical advantages and challenges for IHL

The operational benefit of this tool is clear. It speeds up the process significantly, as well as allowing for more potential COAs to be developed and assessed than would be possible by human planners, which enables the commander to focus on the more creative aspects of planning. Yet by replacing much of the analytical process traditionally performed by the human planners with automated wargaming, the specter of a problem arises in that the decision-making of the process is obscured, and unaccountable. CADET employs a structured knowledge-based system, the most basic of the AI techniques. It wargames various COAs according to predetermined parameters and allows for a human to alter the parameters during and after the process. Even so, it raises the problem of transparency. One of the most important insights from the ICCES experiment came from the users who found that the AI generated synchronisation matrices were difficult to comprehend, despite following the same layout as those generated by human planners. The authors determined that the matrix is only beneficial as a shorthand for recording a human cognitive process, but that without understanding that underlying thought process, it is not as effective.¹²⁴

This is illustrative of the larger problem facing the growing sophistication of AI decision-making systems, whereby the role of the human planner is increasingly diminished as the systems are able to learn and adjust themselves through machine learning techniques.

- Conclusions of the experiment

CADET shows an example of what narrow modular AI can achieve, and signals a likely near future scenario, wherein it forms a part of an integrated set of tools that together replace parts of the manual MDMP. The potential for it to provide enhanced value is considerably boosted with the introduction of more sophisticated AI techniques. The experiment provides some important insights for this future scenario. The authors conclude that the technology used in the experiment is capable of being applied in various conditions, including in the battlefield, as part of an integrated suite of tools and that, with advances in the field of AI more broadly, it may function as a component part of society-wide applications such as long-term strategic processes that involve political, economic, social and ideological concerns.¹²⁵ Finally, operational and tactical battle planning is an example of a real world scenario where distinct problems are closely coupled and are not easily isolated. The tight interleaving technique used in the ICEES experiment is one computational way to solve this kind of problem, but the authors suggest that to reduce technical risks, which are a result of diverse, complex processes, it is better to use simplified algorithms that often trade optimality for speed. This is clearly better for the military application, if speed is the most valuable metric, but will prove problematic for the implementation of other concerns, such as IHL norms.¹²⁶ The trade-off between ‘optimal’ and ‘legal’ AI that arises here will be discussed in Section V.

At this level of its development, CADET is relatively limited in that it is not able to conduct the truly creative aspects of wargaming. While it is able to plan backwards from the high-level COA sketch and statement provided by humans, it is not able to invent any strategy for the battle or take account of overarching strategic considerations. Nor is it capable of taking into account any cognitive or emotional aspects of battle that humans would do instinctively, including the limitations of the humans that would ultimately carry out the plans.

¹²⁴ Rasch, Kott and Forbus (n. 122) p. 24

¹²⁵ A. Kott, and Corpac, P., “*COMPOEX Technology to Assist Leaders in Planning and Executing Campaigns in Complex Operational Environments*,” (2012) 12th International Command and Control Research and Technology Symposium

¹²⁶ General George S Patton is most famously quoted as saying that “*A good plan violently executed now is better than a perfection plan executed next week*”

4.3. How AI works

One consequence of artificial intelligence is that it requires humans to confront how intelligence is understood. Like anything relating to the human composition, human intelligence exists on a spectrum: from being good at basic general functions, like walking without bumping into things, to being experts at highly specialised occupations. It is also the ability to reason well, or to reason at all. There are already AI systems that outperform humans at particular tasks, within specific domains. These are mathematical representations of reality that can receive, process and learn from extremely large and complex sources of information, but are not able to abstract or reflect on the meaning of their tasks, nor are they flexible enough to transfer their abilities across other competencies.¹²⁷ One of the primary goals of AI research is to develop general problem solvers that can emulate humans by receiving high-level information about a particular problem and then automatically determining a solution, or range of solutions, on the basis of on a general world-view that has been constructed from experience.¹²⁸

In discussing the impact of artificial intelligence in any discipline, therefore, it is necessary to distinguish between two separate but related scenarios that the near and long-term future present respectively. Although it is somewhat artificial to separate them, given that modern technology develops along a continuous trajectory and not in discrete pockets of discovery, it nevertheless helps to distinguish philosophical and existential questions about the nature of and ultimate survival of humanity from the more concrete scenarios that law-makers need to address with principles and rules in the meantime. This is in no way underestimates the seriousness of the challenges presented by future AI, but it is thought that by dealing with near-future scenarios effectively, this will have a practical effect on the long-term development of the technology itself, which is currently largely speculative.

The Future: Artificial General Intelligence (AGI)

The long term prospective envisages AI with general intelligence that resembles human intelligence in that it can apply its knowledge and expertise across the full range of different contexts and competencies. An AGI would learn and apply lessons in the way that humans do, across trillions of ‘neural connections’. The argument is that this kind of general intelligence is what would lead to superintelligence in that the AI would be able to outperform humans in specific domains, as it is able to do already in number of domains, and also be able to apply that ‘super knowledge’ to any new problem that it is presented with. Furthermore, an AGI would be able to program itself, meaning that it could adjust or rewrite its own code and generate new algorithms that it determines are necessary for its own improvement which, together with its ability to access all available knowledge via the internet, would lead to an intelligence explosion, or what has been called the singularity. In this scenario humans may be left entirely out of the loop as they would have no control over the AI after the initial algorithm is activated.

This picture is still a distant prospect, and experts disagree on when it might be expected, with estimates ranging anywhere between five and one hundred years. Yet it is important to keep in mind, given that it is precisely this AGI that artificial intelligence research ultimately strives towards.

Present and near-future: Modular AI

Modular AI, on the other hand, refers to applications that are expert in a specific domain, or at completing specific tasks. Modular AI is also capable of learning, and within its particular

¹²⁷ Ayoub and Payne (n. 10) p. 800

¹²⁸ The use of anthropomorphised terminology such as a systems ability to “learn” and “think” is useful for understanding how artificial intelligence works, but is not to be confused with how humans perform these tasks. This is particularly important given that the domain under review is that of human decision-making. Human and machine decision-making are inherently different and must not be too readily equated.

competency it can learn and adapt itself in order to improve its outcomes, but it can only apply its lessons within its specific domain.¹²⁹ Within this broad category of modular AI, which describes all the AI available currently and in the near future, there are degrees of sophistication and techniques that are employed. The advancements taking place are a result of the increasing ability of AIs to learn more effectively. There are, essentially, two distinct levels of development of modular AI.¹³⁰

Expert or rule-based AI. These are systems such as CADET whereby expert knowledge is used to program a computer to produce outputs that have been defined in advance by the programmer. An expert system is provided with the full spectrum of knowledge belonging to a particular domain (or as much of it as possible), and is also given the steps that are needed in order to generate a decision, the basis for making its decision and the rules regarding exceptions. In a rigid expert system, the system will simply follow these steps and this is what is sometimes called ‘pure automation’. A flexible expert system on the other hand will use the information provided and, together with feedback from the programmer, will learn how to generate better decisions by conducting a series of *if-then* statements to infer the best result. In either case, the knowledge base of the expert system can be updated, albeit by a human programmer, to reflect new scenarios.¹³¹ The advantage of expert systems is that they are easy to know and understand, given that it is humans that provide the knowledge base and rules by which it learns. However this technology is considered old generation AI and thus it is machine learning AI that presents the most likely and most challenging future applications of AI.

Machine learning AI. Machine learning is a field of study that uses computer science and predictive statistical techniques to allow an AI system to learn from data by inferring rules or procedures that enable it to predict certain outcomes. It can then use this information to modify or adapt its actions to be better at achieving an outcome without having to be explicitly programmed by a designer or programmer.¹³² This is where recent advances in AI research have been so successful, particularly by leveraging off of increases in available data combined with significantly improved computing power.

Machine learning can either be supervised or unsupervised. Supervised machine learning means that the system is trained by being given a set of examples (inputs) and explicitly told what the desired outcome is (goal). The algorithm learns from the examples and determines a general rule that is required to get from the inputs to the goal. This general rule can then be applied to new inputs and, if successful, still achieve the desired result.¹³³ Unsupervised learning is where the AI is given the inputs but no specified goal. In this case the algorithm is left to identify patterns in the structure of the information and similarities between the inputs on its own, and to derive rules and categories accordingly.¹³⁴

Machines can also learn through reinforcement, which technique lies somewhere between supervised and unsupervised learning. In reinforcement learning, the algorithm is told whether a particular response it produces is wrong, but not what the correct answer is. It then explores possibilities until it figures out the correct answer.¹³⁵ The training data in reinforcement learning is thus only provided as a result of the systems actions. An illustration of this is the AI in a computer game that uses specified rewards such as “win the game” to train the system. The system itself is designed specifically to repeat the task until it maximises its rewards. If the system chooses an action

¹²⁹ Ayoub and Payne (n. 10) p. 795

¹³⁰ U.S. GOA, *Technology Assessment: Artificial Intelligence Emerging Opportunities, Challenges, and Implications*, (2018), Report to the HoR Committee on Science, Space, and Technology, p. 10

¹³¹ Ayoub and Payne (n. 10) p. 795

¹³² Marsland, S., *Machine Learning an Algorithmic Perspective*, (2015), CRC Press p. 5

¹³³ *ibid* p. 6; Brundage and Avin (n. 152)

¹³⁴ Marsland (n. 132) p. 5

¹³⁵ *ibid*

which causes it to be killed, it learns from that consequence that it should not take that action and through trial and error it will discover the best series of actions to “win the game”.

One final method of machine learn is called evolutionary learning, whereby a computer model or algorithm is structured so as to emulate human biological learning processes. One of the most important advances in AI, called artificial neural networks or neural-nets, is derived from this method.¹³⁶ Neural-nets are a family of algorithms inspired by the biological function of the human brain in that they use artificial ‘neurons’ to model complex relationships between inputs and outputs. Associated with neural-nets is the important concept of deep learning, which has given modern AIs extraordinary power and has been the source of most of the publicised AI achievements over the last few years. Essentially, it is a category of algorithms that enable the AI to learn from experience and understand the world in which it operates by creating a hierarchy of concepts. These concepts are not provided by humans but rather are learned from experience, so that the system infers increasingly complex ideas from preceding simpler ideas.¹³⁷ In highly dynamic, uncertain environments where there is no statistical certainty as to how a particular input will actually affect the output, deep neural-nets can be used to find patterns in the input data which the system then uses to adapt to new environments and learn new skills and tactics by updating its underlying structure.¹³⁸ The operational level of warfare is a good example of this kind of environment. There are many factors including different warfighting functions, assets, time frames and targets and it might not be possible to know exactly how each contributes to an overall outcome.

Another example of systems that makes use of evolutionary learning are genetic algorithms, which are designed to approximate, through computation, how evolution works particularly regarding reproduction. Essentially, like biological evolution, the parent algorithm makes genetic alterations that will benefit its derivative algorithms’ fitness depending on its environment.¹³⁹

Finally, there are decision trees which work by creating a predictive model that breaks a problem down into sets of binary choices, also described as ‘if-then’ rules, about each feature starting at the root (the observations or inputs) and progressing to the leaves, (which represent the decision). The system learns by adjusting the results until it discovers an optimal solution. The benefit of these systems are that they are possible to understand.¹⁴⁰ Although much of the advantage to be gained from them, namely their ability to compute vast amounts of data and determine potential solutions much faster than humans, would be lost, it nevertheless is possible to backward engineer them in order to understand how they reached a particular solution. This is unlike neural-nets and genetic algorithms, where their opaque processes result in a ‘black-box’ which is problematic for a number of reasons. The interleaving technique by CADET is derived from the knowledge tree.

This is not a closed list of approaches to machine learning, or AI techniques, but simply a brief introduction to some of the fundamental aspects of artificial intelligence that are necessary to inform a discussion about the potential use of AI in military decision-making in the future, and to know the challenges and opportunities it presents from a regulatory perspective. For example, programmers of current modular AI systems, whether it is a knowledge-based expert system or a supervised machine learning, still play a key role in helping the systems learn and in directing their outcomes. As such, they are at least potentially brought into the realm of accountability for the systems outcomes. Yet understanding how unsupervised machine learning and deep learning neural-nets operate is important in order to recognise the challenges of the future where these systems are mainstream. In that circumstance, where a system that has learned in a way neither predicted nor understood by human-beings, the responsibility of the programmer becomes more difficult to substantiate.

¹³⁶ Marsland (n. 132) p. 43

¹³⁷ Goodfellow, I., et al, *Deep Learning*, (2016), MIT Press, pp. 1-2

¹³⁸ Marsland (n. 132) pp. 43-46

¹³⁹ *ibid* pp. 211-213

¹⁴⁰ *ibid* p. 249

4.4. Incorporating AI into military decision-making: How it might work

Given the current and near future technical capabilities, the next question is how AI tools like CADET, and its more sophisticated future iterations, will affect the decision-making process at the different levels of warfare. The natural world represents different scenarios requiring problem-solving, broadly classified here into three generalised groupings that correlate to the different stages of warfare discussed in Section II. AI works differently in each of these scenarios.

In the first category are those situations wherein the goals are clearly defined and the environment is predictable. This grouping corresponds to tactical level MDMP at its truest form which is concerned with ‘do-or-die’ problems¹⁴¹ (leaving aside the effect that modern warfare has in requiring tactical commanders to engage in more high-level decision-making of a strategical character). In this case the engineering solutions are straightforward, requiring simple automation, the most basic of AI technology. A targeting operation, for example, where the goal is to shoot a particular target, and the weapon, timing and location of the target are clear., could then use a drone that has been programmed to shoot as and when it has a clear vision, according to a range of predetermined criteria. In this case the ‘decision’ to shoot is clearly understandable and traceable to the human operator or programmer of the AI. There are plenty of examples of modular AI in use at the tactical level by the U.S. armed forces, and some that are specifically used to assist with MDMP. At the moment these tools perform narrow functions rather than replacing humans in multiple stages of the planning process in the way CADET envisages, mainly because the cognitive requirements of operational and strategic commanders are fundamentally different from those underlying tactical expertise in ‘do-or-die’ situations. “*Operational problems are characterised by complexity and uncertainty embedded in a turbulent environment riddled with uncertainties.*”¹⁴²

The second category comprises situations where the goals are clear, but the environment is unpredictable so that the means by which the goal is achieved is uncertain. The tasks of AI at this level are comparable to those at the operational level of military decision-making, where commanders are concerned with, among other things, proportionality and military necessity assessments. It takes more sophisticated algorithms than basic automation to engage in determining which lives are more valuable, and deciding on the circumstances under which certain lives can be sacrificed. Continuing the targeting example, the goal at the operational level may be to discover and eliminate a particular target, but there are a number of variables in play in respect of the means and methods of achieving it that need to be coordinated, analysed and planned. A machine learning algorithm, supervised or unsupervised, could in theory be trained to do much, if not all of this work with potentially substantial implications for the division of labour between human and machine. Currently, however, there are practical difficulties in respect of computing this scenario. First, defining what success is and articulating that correctly is challenging. AIs are extremely literal and will do exactly what they told,¹⁴³ sometimes with disastrous consequences, given the impracticality of having to determine all possible iterations of a scenario *ex ante*. They will often take actions that humans could not have anticipated and thus did not account for in providing goals. Such autonomy works well in stable environments but the more unstable the environment the more difficult this will be.

Second, having defined the goals or rewards, the human operators will not be able to predict nor fully understand how the system has arrived at its particular decision because of the inherent black-box of machine learning AI. At best the system may be able to provide an approximate rationalisation of its process. While humans’ thought processes are themselves often opaque, and explanations thereof mostly extremely limited, particularly during warfare, the inscrutability of AI

¹⁴¹ Wallace, R., *Carl von Clausewitz, the Fog-of-War, and the AI Revolution: The Real World Is Not A Game Of Go*, (2018), Computational Intelligence, p. 31

¹⁴² *ibid* p. 31

¹⁴³ Ayoub and Payne (n. 10) p. 807

is at least notionally different. It presents moral as well as legal challenges, the latter largely related to the difficulties in establishing the mental elements required to establish individual liability in international criminal law. Despite these challenges, it is precisely these situations that contemporary and future modular AI will operate best and where AI decision-making systems will likely be most advantageous in the planning of operations that connect tactical implementation to overarching strategic goals, at least for some time to come.

The third category are those situations where both the goals are unclear and the environment is unpredictable. This scenario most resembles the strategic level in that the goals themselves cannot be defined but instead are dynamic and contingent on a shifting, ambiguous and often uncertain interaction between economic, social, political, even religious drivers. Determining what constitutes success under these conditions is a complex and subjective process¹⁴⁴ that does not lend itself easily to computational goals or rewards. Thus AI, at least in its current form, does not function well in this kind of environment as it cannot be trained to get better at something if it does not know what the goals are. This may change with the development of unsupervised learning so that, at least in principle, AI tools will support or ultimately replace strategic decision-making. Yet the technical difficulties related to modelling these environments offers one explanation as to why there has to date been limited use of such tools at the higher levels of MDMP because, in spite of the clear advantages these systems will provide to their users, there has been little attempt to pursue their development by the military establishment, particular so higher up the echelons of the command structure.

Advantages and Challenges

As the authors of the ICCES experiment pointed out, the area of MDMP is ripe for incorporation of intelligent automated systems, if not yet to replace them, then to take over large parts of the process.¹⁴⁵ The manual process in operational planning, for example, is cumbersome, slow, and inflexible, often working off of imperfect information and under significant time constraints. The success of a plan is very reliant on the synchronisation of various units and assets that are independent but inter-dependent, the specific concentration of firepower, all coming together at precise moments. The quality of the output will be largely dependent on the size of the staff as well as their knowledge, experience, how well they work together as a team, and how good they are at estimating outcomes. Notably, it will also be subject to the nature, biases and other cognitive shortcuts of the individuals involved.¹⁴⁶ This all in relation to serious decisions that affect life and death. AI systems offer significant advantages in this regard, ranging from the purely tactical to broader advantages that address some of the shortcomings of human decision-making. These benefits are, of course, accompanied by certain challenges, particularly in respect of regulatory oversight.

- Military and cognitive advantages

The advantages that modular AI systems offer will depend on the specific technique employed, and these are better suited depending on the specific features of the different scenarios as described earlier in this section. In general, however, they are able to outperform humans in particular domains because of two primary characteristics. The first is their ability to process information in parallel. While a human planner can consider one or two solutions at a time, AIs can consider thousands, depending on the processing power of the system, while avoiding the problems that arise in humans from a data overload. Also, despite being modelled to mimic human cognitive processes in some cases, they are intrinsically different to a human brain, and thus AIs are often successful at identifying patterns and solutions previously invisible to humans.¹⁴⁷ The second characteristic is

¹⁴⁴ Ayoub and Payne (n. 10) p. 811

¹⁴⁵ Rasch, Kott and Forbus (n. 122) p. 19

¹⁴⁶ Kott, Budd and Ground (n. 116) p. 3

¹⁴⁷ Ayoub and Payne (n. 10) p. 807

that, being inorganic, they are free from the biological shortcomings that impede humans, including fatigue, stress, a need to spend time consuming energy sources, and a tendency to make errors resulting from boredom, distraction and repetition. In respect of military applications, this potentially offers a significant tactical advantage both in terms of saving time and outwitting an adversary with unexpected maneuvers. The example of CADET reveals how the system uses considerably less person-hours in order to achieve comparable results. Machine learning will probably reduce this even further, but its primary benefit is its ability to achieve much better results, depending on the nature of the goals that the AI is used to achieve. They will be free of a number of the cognitive limitations natural to human brains. For example, AIs will avoid assumptions that arise from emotions, ego and heuristics. One possibility of machine learning is that, even if human biases are encoded into their original structures (intentionally or not) AIs will not be bound to entrenched cognitive ideas or ‘groupthink’ the way that humans are. Instead they will learn and adapt to new behaviours that are better aligned with their ultimate goals, thus potentially being able in the future to challenge previously held assumptions, and to critically examine the underlying rationale for engaging in a particular conflict.¹⁴⁸ Some authors argue that as a result of these advantages, AIs will be more successful than humans at reducing civilian casualties and that, if this is the case, there is a moral imperative for their adoption into military processes.¹⁴⁹

In the same way that AIs have been able to outperform the world champions in extremely complex strategy games,¹⁵⁰ adversarial planning and scheduling applied to military decision-making would render the output much more difficult if not impossible for adversaries to predict or respond to as quickly as the AI. It is their ability to process vast quantities of data, to predict and react considerably faster than their human counterparts, to reason independently of emotion (for better or worse) and avoid unethical human impulses and cognitive shortcuts and, finally their political disinterestedness that enable them to outperform human planners in increasingly complex situations. Applied to operational level decision-making, these advantages will all be aimed at planning maneuvers faster, more efficiently and unpredictably for adversaries. However, if they are able to outperform human adversaries then they are almost guaranteed to also be outperforming their human operators either in their speed or their ability to comprehend the systems cognitive processes. This will require a trade-off between the benefits that such systems could provide operationally, and the need for responsible and legal AI.

Finally, the crucial ability of the different levels of warfare to understand one another’s goals and decisions, and to communicate effectively, will be enhanced by the use of AI, particularly when, as previously mentioned, the Internet of Things is incorporated into the military domain and the systems are able to communicate directly with one another. The greatest disasters in warfare often stem from a failure of communication between these different levels of warfare. As Mao Tse-Tung wrote in 1936:

“Why is it necessary for the commander of a campaign or a tactical operation to understand the laws of strategy to some degree? Because an understanding of the whole facilitates the handling of the part, and because the part is subordinate to the whole. The view that strategic victory is determined by tactical successes alone is wrong because it overlooks the fact that victory or defeat in a war is first and foremost a question of whether the situation as a whole and its various stages are properly taken into account. If there are serious defects or mistakes in taking the situation as a whole and its various stages into account, the war is sure to be lost.”¹⁵¹

¹⁴⁸ Ayoub and Payne (n. 10) p. 807-808

¹⁴⁹ Arkin, R., *Warfighting Robots Could Reduce Civilian Casualties, So Calling for a Ban Now is Premature*, (2015), IEEE Spectrum

¹⁵⁰ DeepMind (n. 5)

¹⁵¹ Tse-Tung, M., *Selected Military Writings of Mao Tse-Tung*, (1963), PRC, pp. 79-80

And yet the ability of AI systems to communicate rapidly among themselves, with or without human oversight, is just one of the characteristics of AI in military decision-making that is both potentially advantageous and presenting of hard challenges.

- Technical challenges

The challenges that AI presents are largely connected to the capabilities of the technology and the seriousness of outsourcing life and death decisions. Probably the most significant challenge, as mentioned above, is the problem of the black-box, the inability of humans to see into or understand the cognitive process of machine learning AI systems. Their ability to learn in a way fundamentally different from human brains means that it may never be fully understood.

There are at least two areas of AI research that aim to address this problem. One is seeking a way for machines to record and understand their own processes, and the other is concerned with enabling such systems to communicate effectively with its users. The potential advantages of human-machine interfaces can only be fully realised if researchers are successful at cracking this problem of communication. The same goes for any regulation that requires having a ‘human on the loop’. If humans are to remain accountable for the decisions that result from this collaboration, they will have to be able to understand how such decisions were reached. Another technical challenge, called Goodhart’s Law, essentially refers to the fact that once a metric becomes a goal in itself, it ceases to be a good metric. This applies beyond AI, but in relation to AI systems, it describes how, as they become get increasingly intelligent, AI gets better at finding tricks and hacks that enable them to maximise their reward without concern for the disruption it may cause to their environment, or without necessarily meeting the ultimate human objective.¹⁵² The implications for military planning are substantial. One final challenge arises essentially from the reverse of the advantage it provides in respect of avoiding human cognitive shortcomings. As much as a system may not have impulses of rage and revenge, it will also lack the better angels of human nature, such as bravery, compassion, and self-sacrifice, which also reveal themselves in times of war and which are different from legal, and even ethical behavioural parameters. Again, this may be able to be incorporated into its design by a programmer, but it is not yet certain whether (i) this will be possible technologically, given that there are no apparent rules for such behavior, even in humans (these kinds of acts often take us by surprise) and (ii) even where these have been written into the AI’s ontology, whether machine learning will mean that the AI sees these behaviours as a defect, and ‘unlearns’ them.

Examples

Given the state of AI development to date, there is not yet an AI that can replace the human decision maker at the operational and strategic levels, even if there were the political or organisational will to use it. Nevertheless there are a number of ways in which modular AI is able, at least theoretically to assist such decision-makers through human-machine cooperation.

- for developing basic strategies¹⁵³

One example of modular AI being used for developing strategy is DeepMind’s multi-use, deep reinforcement learning algorithm that uses reinforcement learning to enable it to play 8-bit Atari games and maximise its ability to get the highest score of the game (its reward). The machine quickly has surpassed expert human level.¹⁵⁴ This shows how such AI can have a major impact on other spheres of strategical decision-making (although in a limited domain here), but especially important is that researches found it difficult to fully understand the strategy that the algorithm employed. The only way they could do so in fact was to infer the strategy from the tactical actions the system

¹⁵² Brundage, M., and Avin, S., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, (2018), FHI

¹⁵³ Ayoub and Payne (n. 10) p. 802

¹⁵⁴ Mnih, V., et al., *Human-level control through deep reinforcement learning*, (2015), Nature, p. 529-533

executed¹⁵⁵ (i.e. by looking backwards) which is neither certain nor efficient in the context of military strategy.

- for assessing risk

Some modular AI is able to quantify risks and provide clear bounds on the probability distribution of that risk by providing a decision-maker with the probability of a desired outcome, together with its certainty (expressed by a percentage) of that probability. This would be expressed along the lines of an AI reporting that it is 95 percent sure that a particular action has an 85 percent chance of reaching the desired outcome. This quantification of risk can then be factored into the decision by the decision-maker.¹⁵⁶ Under ordinary decision-making processes, military commanders must make this risk assessment by themselves. Even accounting for the fact that good commanders should be better at this calculation than others, humans are still notoriously bad at this as a result of being at the hand of imperfect information and a range of other limitations, and there AIs can add value during this exercise.

- Visualisation

One of the biggest areas of progress of modular AI to date has been on image recognition and classification. This is a big part of the controversial Project Maven,¹⁵⁷ as it is a fundamental aspect of building an autonomous weapon system including non-lethal systems. With current machine learning techniques, enhanced by deep learning, AIs have already been successful in extracting certain features, and using them to classify objects.¹⁵⁸ The implication of this for MDMP is important. For example, IHL provides relatively clear parameters regarding what constitutes a legitimate target and by learning the features of such targets an AI could be able to recognise new targets and conversely recognise what cannot be targeted. Again, this may have potential benefits for IHL and military efficiency if it is better at the assessment than humans.

- Microworlds

‘Microworlds’ are abstract representations of the battlefield environment that use reinforcement learning to conceptualise the terrain, force deployment, and enemy movements and responses. They are essentially a simulated wargaming process that has the potential to outperform humans, not only in terms of speed but also in making recommendations that human planners would not have anticipated.¹⁵⁹ These simulations are particularly useful for higher level strategic and operational decision-making in that they can access and analyse immense quantities of data and thus recognise patterns that a human could never have done, which in turn challenge the heuristic assumptions that have driven strategic planning throughout history.

¹⁵⁵ Ayoub and Payne (n. 10) p. 803

¹⁵⁶ Ayoub and Payne (n. 10) p. 804

¹⁵⁷ Officially called the Algorithmic Warfare Cross-Functional Team, Project Maven was established by the U.S. DoD in 2017 to accelerate the integration of AI and machine learning across its operations. For more information see United States Department of Defense, *Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)*, 26 April 2017

<https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf>

¹⁵⁸ Ayoub and Payne (n. 10) p. 804

¹⁵⁹ Scholz, J., *et al.*, *Machine learning for adversarial agent microworlds*, (2005) MODSIM, pp. 2195-2201

4.5. Operationalisation of AI decision-making into the military

Given these potential advantages, and the quite advanced stages of development of decision-making, planning and strategy tools in the private sector, military use of AI in aiding its decision-making processes is surprisingly minimal. There are some tools that have been operationalised, but these are limited to conducting specific, low-level tasks such as analysing battlefield terrain and recording decisions, and are not cutting edge technology. More advanced applications in use by the U.S. armed forces are limited to tactical level use of force¹⁶⁰ and intelligence gathering, including the systems employed by the NSA to mine large amounts of data and recognise patterns using machine learning.¹⁶¹ There are a few possible reasons for this lack of operationalisation of the technology into MDMP at the higher levels. One is that the development of the technology simply has not yet reached a level of proficiency to confer the perceived advantages, or at least that it has not demonstrated its abilities sufficiently to establish trust in human users. Another is that any research into this particular application of AI stalls because of the difficulties associated with training and testing these systems in a meaningful context. They would not be able to be trained in a genuine war situation as the practical and legal ramifications would be serious.

A third reason stems from a more fundamental reluctance to hand over the heretofore most creative aspects of military decision-making, the art of war, which has historically depended on the experience of a commander and his expertise and brilliance in respect of drawing together many moving parts and applying them to war theory, history and with pragmatism. The largely subjective exercise of military planning is a highly creative undertaking, combining qualitative and quantitative assessments. It involves anticipating and exploiting the weakness of one's adversary, the environment and the strengths of one's own troops, in an effort to bring about the strategic goals through the precise planning of specific operations all at once, often in highly demanding situations where information is imperfect. It is therefore unsurprising if there is an institutional distrust in the ability of AI systems to accurately capture and understand the subjective motivations that drive military decision-making, and which increase as decisions take on strategic character. This skepticism is especially plausible in respect of the innovative and relatively unknown machine learning, which will inevitably fail while it is developing.

As a result, no such tools currently exist that would remove the human decision-maker altogether at the strategic and operational levels. However, the combination of increased pressures of combat and the growing sophistication of deep learning has meant that there has been significantly more adoption of AI at the tactical level,¹⁶² and there are a number of ways this influences operational and strategic decision-making, albeit indirectly. For operational commanders, the availability of advanced technology becomes a significant factor in the planning of operations as they must determine whether their use is proportionate to the military objective, and whether they are in keeping with the overall strategic goals of the armed conflict (namely to have a quick but harsh conflict, or a drawn out conflict aimed at wearing out the adversary by attrition). For the strategic level, access to AI technology at the lower levels may influence the decision whether or not to resort to use of force at all, or whether to pursue diplomatic or other means instead. Some critics of AI argue that it lowers the barriers to entry, particularly in respect of political support of populations, given the reduced risk to the lives of the armed forces. In a conflict in which both sides have access to the technology this could potentially be a positive development but the greater likelihood, given the nature of international relations, is that it leads to increasingly asymmetric warfare. Another consideration for strategic and operational decision-makers will thus be whether or not the adversary

¹⁶⁰ Some examples are the Aegis combat system that uses computer and radar technology to track and guide weapons towards enemy targets, and the MIM-104 Patriot, a surface-to-air missile and anti-ballistic missile system that intercepts missiles using high-performance aerial and radar systems

¹⁶¹ Hickens, M., *How the NSA could get smart so fast*, The Wall Street Journal, 12 June 2013

¹⁶² Ayoub and Payne (n. 10) p. 806-807

has access to AI systems at the tactical level. For operational planners their choices will be made for them if the answer is in the affirmative.

Another way that the influence of AI at the tactical level works its way up to the operational and strategic levels is by advances in the technology itself. The authors Ayoub and Payne use the example of the U.S. Navy's X-47, an experimental unmanned aerial vehicle platform, which were originally designed to be autonomous but currently function as expert systems commanded to carry out specific, low-level tasks, such as flying to a particular point or shooting at a particular target. There is still considerable input from humans in their overall operation. As the technology advances however, these instructions can change to more high-level ones such as, "protect the carrier", leaving decisions regarding the intermediary steps to the AI.¹⁶³ This way operational decisions creep into a system initially accepted as making only limited tactical decision-making. Furthermore they suggest that *"in the future, [similar] decisions about the identity of possible adversaries will be increasingly automated, not least since opposing systems will otherwise be technologically able to thwart human operators. An attack by swarms of agile, autonomous actors can only be defended with systems that operate quickly, with autonomy and intelligence."*¹⁶⁴ This will have significant implications for IHL principles of distinction and proportionality.

The assumption, particularly in public statements made by members of the U.S. military, is often that machines will increasingly take over basic tasks, while the human will retain the creative function of warfare. However, given the sophistication of the technology and its tendency to creep upwards, it is necessary to retain some skepticism. CADET exposes what even fairly simple AI techniques can achieve in respect of replacing human planners in the process of wargaming and may prove to be an early iteration of what Ayoub and Payne call an 'oracle-like' AI assistants. They predict that strategic decisions will remain with humans for some time to come but will be supported by these AI oracles that provide recommendations. While they leave the question there, it is for international lawmakers to take this further, in considering whether this will affect the implementation IHL norms.

Finally, the current attitude of the U.S. military establishment regarding strategic decision-making, while somewhat comforting, may change as a result of shifting international power dynamics, political threats of potential adversaries and the operational advantages of increasingly sophisticated technology. In addition, the U.S. reluctance to remove the decision-making process from the established chain of command, is not necessarily a guarantee that others will do the same.¹⁶⁵ International law should establish its principles and parameters with which to address this scenario before state practice and ubiquity of the technology renders it an even greater challenge.

4.6. Conclusion

Although AGI is some way off, modular AI is being applied to increasingly more areas of life and with increasing sophistication, including some applications that are highly relevant to strategic decision-making. In the military context such AI promises to profoundly increase the tactical advantage of those who use it, and potentially operational and strategic decision-making, at least insofar as it is faster, more efficient and avoids some long-standing human cognitive shortcomings.¹⁶⁶ These advantages, however, are to be held up against the shortcoming of modular AI that cannot discern the subjective meanings of considerations of strategy, such as to "break the enemy's resolve" or "secure order and justice", terms that have been described as "unknown unknowns".¹⁶⁷ Even with access to all the information in the world, the complexity of human

¹⁶³ *ibid*

¹⁶⁴ *ibid*

¹⁶⁵ Ellman J., *et. al.*, *Assessing the Third Offset Strategy*, (2017), CSIS, p. 13

¹⁶⁶ Ayoub and Payne (n. 10) p. 807

¹⁶⁷ Donald Rumsfeld, *Department of Defense News Briefing*, (2012) available at <http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636>

relations means it may never be able to predict strategic outcomes with certainty. It is curious that, although humans are largely subject to similar shortcomings in this regard, the expectation of certainty or perfection are significantly increased when AI is brought into the equation, probably because of a moral abhorrence of machines making life and death decisions. According to some experts, therefore, “*strategy that involves humans, no matter that they are assisted by modular AI and fight using legions of autonomous robots, will retain its inevitable human flavor.*”¹⁶⁸

The attitude of the U.S. military establishment, at least for now, appears to conform to these overall concerns. Instead DARPA has concentrated its focus on research efforts aimed at developing supporting technology, such as increasing the capacity for communication between the humans and AI systems, enhancing the exchange of information, and addressing the black box.¹⁶⁹ Thus at the higher levels of decision-making, this is a kind of human-machine sharing of tasks – where there is a human on the loop – is currently weighted in favour of humans conducting the majority of the process, with AI performing discrete, clearly defined tasks at the tactical. However, improved AI technology, combined with growing political competition potentially sparking a new arms race, and increasing deployment of tactical AI, will inevitably drive this towards a shift in the equation whereby distribution of tasks becomes increasingly weighted in favour of intelligent systems.

In light of the potential future paradigm wherein the incorporation of AI processes into military decision-making will bring advantages to each level of warfare, it becomes necessary to consider the effect it will have on the implementation of IHL, and how it can address any challenges to its implementation. If the challenges presented by this scenario are *sui generis*, then the international community may need to establish an entirely new set of rules by which to govern it. If, on the other hand, the problems presented are merely a new, intensified manifestation of existing problems, then it is possible to look to existing frameworks to safeguard the implementation of IHL during the military decision-making process. The doctrine of superior responsibility is one such avenue that could address the future scenario of AI assisted decision-making.

¹⁶⁸ Ayoub and Payne (n. 10) p. 816

¹⁶⁹ See for example DARPA, *Explainable Artificial Intelligence (XAI)* available at <https://www.darpa.mil/program/explainable-artificial-intelligence>

SECTION V – FUTURE PARADIGM: AI, MDMP AND IHL

The previous section explored how AI decision-making tools might affect the MDMP, whereby much of the process currently performed by humans will be transferred to the AI system, more and more so as they gain in technological sophistication. Some of these tasks will be more easily transferred, others less so, from both a practical and a legal perspective. The question of how it will affect the implementation of IHL depends first on the sophistication of the technology, the specific AI techniques that are employed to train it, and the availability of enough good quality data; second on the nature of the decision or planning processes being conducted; and third on the ideology of the user. AI, like many technologies, is user-sensitive, and in the wrong hands is subject to misuse, but that alone does not mean the problem is insurmountable. The purpose of the law is to establish what would constitute misuse in respect of these tools and to address those cases where it is improperly deployed. This section is a discussion of superior responsibility as providing the necessary framework by which to establish what constitutes proper use, by relying on the ‘softer’ principles of responsible command. Finally, it looks at the question of individual liability under the new regime.

However, it is necessary first to briefly discuss another potential avenue, namely a conventional review under Article 36 of Additional Protocol I, and to explain why it is not the focus of this thesis.

5.1. Article 36 Weapons Review

One possibility for addressing AI decision-making systems is to categorise them as a means of warfare under Article 36 and therefore subject to review according to the assorted treaty and customary rules applicable to weapons, means and methods of warfare and the principles of humanity and the dictates of public conscience.¹⁷⁰

As noted, combatants’ right to choose their means methods of warfare is limited by some fundamental IHL rules regarding the conduct of hostilities found in Additional Protocol I, together with other treaties that prohibit the use of specific weapons, and in customary IHL. The idea of reviewing the legality of new weapons technologies was first incorporated into an international instrument in the St Petersburg Declaration of 1868. The rule is now captured in Article 36 of the Additional Protocol I which requires that all new weapons, means and methods,¹⁷¹ be assessed and reviewed before they are developed, acquired and introduced by the military for use in armed conflict, to ensure that IHL does not prohibit it. It thus aims to prohibit, or impose restrictions on, the use of weapons, means and methods that would violate international law by determining their lawfulness before they are developed, acquired or used by states. This rule applies to all states, whether or not they have ratified Additional Protocol I, because of the critical role it plays in ensuring that the prohibition from using weapons, means and methods unlawfully is respected in all situations of armed conflicts.¹⁷² The ICRC notes the importance of this review procedure, especially in light of the pace of development of new technologies today. Together with the requirement of Article 82 of Additional Protocol I that legal advisors be available at all time to advise commanders on IHL and “*on the appropriate instruction to be given to the armed forces on this subject,*” these two provisions constitute a framework for ensuring that armed forces conduct hostilities in accordance with IHL, through legal review of planned means and methods of warfare.¹⁷³

¹⁷⁰ ICRC, *Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, (2006) p. 939

¹⁷¹ “Means and methods of warfare” refers to the tools of war and the ways they are used

¹⁷² ICRC, *Guide to the Legal Review of New Weapons, Means and Methods of Warfare*, (n. 170) p. 933

¹⁷³ *ibid* p. 934

One possibility, therefore, would be to argue for a total ban on their use, along the lines of the arguments being made in respect of LAWS. This, however, is problematic for several reasons. The first is because of the foundational nature of AI technology. AI technology has been described as a foundational in the way that electricity is foundational. Both give rise to a myriad of applications, some dangerous and others that have such profound potential to enhance the quality of human life that arguments about banning it entirely are rendered implausible. Another reason relates to the difference between these systems, and LAWS. Arguments for banning the development and use of LAWS are grounded in part (although not solely) on an implicit recognition that lethal decisions ought not to be made and carried out by an autonomous system without human intervention. Activists are therefore concerned with securing that humans also maintain meaningful human control over the use of lethal robots in the field of warfare.¹⁷⁴ As this argument is about lethal weapon systems, it fits comfortably within the purview of Article 36.¹⁷⁵ Nevertheless, they have consequences for the implementation of IHL, besides being a precursor to LAWS, and thus require a framework of regulation.

According to the ICRC's commentary on Article 36 'methods and means' include weapons in the widest sense, as well as the way they are used. Some weapons, such as poison, are therefore unlawful in themselves, while others are not per se unlawful but are subject to abuse. An example of the latter is a precision weapon which is generally permitted but if directed against a civilian population would be unlawful.¹⁷⁶ In its guide on weapons review, the ICRC defines the material scope of Article 36 broadly to include, among other categories: (i) non-lethal weapons systems, (ii) the ways that weapons systems are used pursuant to military doctrine, tactics and so on, and (iii) whether acquired pursuant to research or 'off-the-shelf.' States, however, have interpreted the definitions more restrictively. In the U.S., "weapons" refers to "*all arms, munitions, material, instruments, mechanisms, or devices that have an intended effect of injuring, damaging, destroying or disabling personnel or property.*" "Weapon systems" refers to "*the weapon itself and those components required for its operation, including new, advanced or emerging technologies which may lead to development of weapons or weapon systems and which have significant legal and policy implications. Weapons systems are limited to those components or technologies having direct injurious or damaging effect on people or property (including all munitions and technologies such as projectiles, small arms, mines, explosives, and all other devices and technologies that are physically destructive or injury producing)*".¹⁷⁷ These definitions are typical of other nations view on weapons and weapons systems. In the face of these limited definitions of "weapons" and "weapons systems" that most states have adopted in their rules governing Article 36 reviews, it is complicated to make the argument that AI decision-making systems constitute a means of warfare.

Consequently, AI decision-making systems, unlike lethal autonomous weapons, are not strictly speaking weapons systems. Although they present similar problems in respect of distinction, proportionality and accountability, there is a sufficient gap between their output and the execution of lethal force that makes it difficult to ground an argument for banning them on existing international law. However, their use becomes more widespread in the future there may be sufficient justification to argue for an expansion of these definitions to include other autonomous applications such as AI decision-making systems. Besides being able to assess and recommend the means and methods of warfare to commanders, they would then come to be a means of warfare themselves, and thereby subject to restrictions as to when and how they can be used.

¹⁷⁴ See generally Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (2012)

¹⁷⁵ Hereafter referred to as Additional Protocol I, which requires that "In the study, development, acquisition or adoption of a new weapon, means or method of war, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."

¹⁷⁶ *ibid* p. 936

¹⁷⁷ Hays Parks, W., Office of The Judge Advocate General of the Army, *Weapons Review Programme of the United States*, (2001)

The advantage this route offers is that it would apply generally across all levels of decision-making. There is a conceptual difficulty that arises, however. Under contemporary military doctrine, there is an abstract separation between the intention of the commander and the means of warfare used to achieve it which is evident in the construction of the MDMP. By classifying AI decision-making systems as a means of warfare themselves, there is a real risk that this separation will disappear, which is a subject for further exploration. The Article 36 approach is related to superior responsibility in that they both involve a decision about whether and how to use AI decision-making systems for MDMP, as a means of warfare. The focus of this thesis, however, is on augmenting and relying on the international standards of conduct of military superiors under the doctrine of superior responsibility and the principle of responsible command. In some ways it involves a more nuanced exercise as the criteria will differ for tactical, operational or strategic decision-making. Still, it avoids the conceptual difficulties that arise in respect of conflating the commander's intent with the means of warfare used to achieve it.

5.2. Superior Responsibility: minimum standards

Section III discussed the importance of superior responsibility as a set of minimum standards aimed at preventing violations of IHL, particularly in respect of the principle of responsible command, and the duties of commanders during the decision-making process at the operational level. In truth, however, this area of the doctrine has not received much attention, either in practice or from legal scholars. The emphasis is primarily on superior responsibility as a means of external, *ex post facto*, enforcement through the individual criminal liability of superiors. Despite this lack of attention, the principle of responsible command has a fundamental logic that could underpin the amplification of this aspect of the doctrine, beginning with minimum standards and advancing towards hard law through interpretation or ultimately even convention rules. A treaty, for example, that would oblige states to incorporate rules on the use of AI decision-making systems into their military doctrine, rules of engagement and legislation, is conceivable. The logic lies in the duty of superiors to prevent violations of IHL throughout the conduct of operations, by maintaining awareness and control commensurate with their level of command. The most immediate and tangible impact AI decision-making systems will have on the implementation of IHL, however, may in fact be that they undermine the role of legal advisors in the decision-making process, critically at the operational level of MDMP.

The legal advisor

By requiring significant involvement of military advisors at each step of the MDMP, current U.S. military doctrine indicates an ideological alignment with the aims of superior responsibility to imbue the whole conduct of armed conflict with respect for the law. However, the increasing ability of AI to outperform humans in speed and quality of decision-making raises the question of what the role for the human decision-maker and legal advisor will be in the process of MDMP in the future and whether future military doctrine will maintain this trajectory towards increasing alignment with IHL.

It is apparent that the involvement of legal advisors differs at each stage of warfare, depending on the complexity of the decisions and planning involved. At the tactical level, where decisions are more confined to specific actions, U.S. policy does not require extensive involvement of JAs. Legal advisors are most necessary at the strategic level, given the complexity of decision taken there, but the upshot of that complexity is that their role in strategic MDMP is not considered to be under significant threat for the time being. The primary focus is therefore on operational level decision-making, where near-future AI will offer the most potentially game-changing advances and where the active role of legal advisors is most threatened.

The analysis of the JAs' role revealed the importance of participation throughout MDMP precisely because each stage involves legal rules and obligations of which a superior must be made aware. If even a basic AI tool like CADET is able to receive the high-level mission intent and sufficient battlefield information to create and wargame COAs without any human intervention, there are

already a number of MDMP steps taking place without counsel from a legal advisor. In practice, then, legal advisors would be relegated to merely reviewing plans after they have been made. Reviewing a plan only at the end of the process means that the legal advisor will not be able to fully comprehend the complex assessments that have taken place and will, therefore, struggle to advise properly on the legality of an operation. In these conditions the legal advisor is hamstrung, which position would be exacerbated by the fact that, unlike humans, AI systems are unable to explain their cognitive processes, at least not at the current state of the technology. A person may be able to infer its assessments by reverse engineering (back-propagating) the system but this would never be entirely accurate, and it would undermine the advantages of speed and ingenuity offered by such systems. This goes to the heart of the challenge of AI decision-making systems. Even if they are able to take account of IHL considerations, and to conduct the subjective, quantitative proportionality and military necessity assessments, the cognitive process is shrouded in an opaque black-box that is inaccessible to the human planner. Is he or she then expected to trust blindly in the AI?

Besides requiring that legal advisors be available when necessary and at the appropriate level, IHL, including the rules applicable under superior responsibility, is not overly prescriptive in respect of explicitly demanding legal advisors during military decision-making. However, the need for such advisors is implicit in the obligation of superiors to comply and ensure compliance with IHL by subordinates. The core function of the legal advisor, and the rationale behind the increased role in MDMP, is to ensure that an operational commander has a proper understanding of the often complex legal implications of proposed actions or battle plans prior to, and during, their execution. The commander will probably not be a legal expert him or herself, and hence the importance of the legal advisor. This rationale must be sustained regardless of new technologies, and in fact, should inform any inclusion of such technology into the decision-making process. The intention is to avoid blind incorporation of systems that will disrupt a seemingly positive trend towards expanding the role of the legal advisor and generally greater compliance with IHL during planning processes. Minimum standards of responsible command would thus forbid the use of AI decision-making systems that inhibit a commander's ability to fully appreciate the legal implications of any proposed COA. The reduced role of the legal advisor during the MDMP would be a red-flag in this regard, insofar as it further inhibits the commander's ability to properly comprehend the legal implications of a proposed COA. Furthermore, it may be that this reduced role constitutes a contravention of Article 82 of Additional Protocol I, in that legal advisors could no longer be said to be operating at the appropriate level. Whether this will in fact be the case depends on how much of the task is handed over, and how capable the systems are of incorporating IHL norms, but the use of any tool that leads to a direct contravention of an explicit rule of IHL would not fulfil the principles of responsible command. The critical thing is to ensure that the additional fog of AI does not have the effect of subsuming the (relatively) successful negotiation between the respective logics of military success and IHL that has resulted in the involvement of legal advisors in the MDMP.

The commander

Commanders' responsibility, however, is much broader than listening to legal advisors. Besides being aware of the legal implication of their decisions, they are required to retain control over the final orders and to ensure that they are capable of being carried out within the law. The specific rules and standards are determined by the level at which MDMP is conducted, and the size and nature of the decision, some of which lend themselves better to computation than others. The narrower and more defined the rule, the easier it is translated into a parameter or feedback for a modular AI, whether it uses machine learning or not. That explains why tactical actions are more easily translated into AI systems and have seen greater operationalisation into military applications. As the assessments become more subjective, corresponding to the movement up the decision-making hierarchy, so the difficulties in computing IHL norms increase.

One of an operational commander's first duties in MDMP is to use as much informational intelligence as possible to get the fullest understanding of the tactical situation. AI promises to

enhance this aspect of the process significantly. Through a combination of image recognition and classification, navigation and other advances in the technology, AIs will be able to create a far more detailed picture of the battlefield including identifying objects marked with the protective emblems, and establishing the location of protected buildings and objects, and the concentration of civilians in particular areas. These same applications of AI will be able to contribute to MDMP in finding alternative routes for attack and evacuation of wounded persons and POWs, and finding passages for humanitarian relief. Besides the tactical advantages, this is an example whereby, if used properly, AI could enable even better compliance with IHL than humans could obtain through the accuracy and breadth of options it could suggest, and by reducing the risk of mistakes that result from misinformation.

The possibility for even better compliance with the law comes from the fact that, under the principles of distinction, proportionality and precaution, IHL offers some reasonably clear rules in respect of distinguishing categories of persons and objects, and dictating who can be targeted and under what conditions. Binary norms, such as the prohibition of an armed attack against items marked with the internationally protected symbols, lend themselves quite well to being translated into algorithmic protocols, possibly even more so than in some non-military applications. Consider the famous problem of the self-driving car having to choose between the life of a pedestrian child or the car's passenger. A reason that this question is so controversial is that it forces one to submit the value of different persons' lives to a hierarchy in an *ex ante*, calculated, deterministic way. In the military context, however, this hierarchical exercise has been done and codified, precisely because it is a context defined by violence and life-and-death decisions. A key issue, however, lies in the qualification 'if used properly,' which hints at dual-purpose nature of AI technology. However, again, the role of the minimum standards is to establish what 'properly' means in the context of MDMP.

The greater challenge, however, arises in respect of the more subjective assessments that commanders are required to conduct, and where compliance or non-compliance exist on a spectrum, rather than a simple yes-or-no. Operational commanders are required to make numerous difficult assessments during MDMP such as: what constitutes 'feasible measures' to spare civilian lives and objects during an attack taking into account military and humanitarian considerations? Is the expected collateral damage proportionate to the military objective of an operation or is it excessive? What constitutes a 'sufficient military advantage'? There are also those limited situations wherein a commander is required to conduct additional assessments about the legitimacy of a particular action based on the tactical situation on the ground, or where military necessity permits a derogation or waiver of a particular IHL rule. All of these various judgment exercises are further dependent on the nature of the operation and whether it is a pre-planned or an improvised action.

Theoretically, machine learning AI systems will be able to conduct these complex and subjective assessments in the future if they are not already able, but how successfully they do so will depend in part on the quality of the data used to train the system, and what the programmer and the system itself identifies as 'success' in a given situation. This second factor is one of the most challenging things about AI generally. Depending on certain variables, an AI decision-making system may determine 'success' as anything from total victory, limited victory within the bounds of IHL, complete compliance with IHL (even if that leads to military defeat), or anything in between. Goodhart's law reveals the problems that occur when goals are poorly described, or do not take account of how the rewards used to guide a system towards a goal themselves, become the goal of the AI. If for example, a future version of CADET that uses machine learning is given an operational objective that has been too narrowly defined, this could easily result in a plan that, if executed, would be considered indiscriminate, i.e. would fail a proportionality assessment. The challenge for lawmakers will be to establish clear standards requiring that compliance with IHL be built into systems' notion of success in a responsible and reliable way. They will also need to establish superiors' duty to ensure that this is the case, based on the principle of responsible command, and the consequences for failing to do so.

Another challenge lies in understanding the cognitive processes of the system. This is an area which both private AI researchers and DARPA are paying attention to in trying to discover different ways that AI systems will be able to identify and communicate the salient and mutable features of its own cognitive process. The DARPA XAI (explainable AI) programme, for example, is one of several that aim to enable a ‘third-wave AI system’ whereby machines can understand their context and environment and build models that can explain themselves to humans in an understandable way. A primary goal of the programme is to create trust, and to ‘effectively manage the emerging generation of artificially intelligent partners.’¹⁷⁸ These are just some of the obstacles facing AI experts and engineers that fall under the banner of inquiry known as ‘value alignment’. Value alignment is a key area of AI research which looks at finding ways to ensure that AI systems can recognise and align their behaviour with human values.¹⁷⁹ The demand for value alignment forces a trade-off between the ability of the system to optimise its output, in this case to rapidly determine a COA with the best chance of a military victory, and the control that humans maintain over it to ensure that it achieves this without violating IHL rules. By definition, the more human oversight is required over the system, the more its value is reduced. While this may be inadequate for military-minded persons, it should be noted that this is precisely the compromise underlying the logic of IHL generally, which functions as a limitation on the military normativity that seeks victory through force, by imposing a cost on unrestrained use of that force in the name of humanity. Artificial intelligence adds another layer to this compromise, and just as military necessity is the point of intersection between military and IHL norms, the notion of value alignment is the intersection between the military, IHL and AI.

Some proponents of AI have argued that it will be possible in the future to incorporate human values, including legal rules, into the underlying structure of any AI system. Generally, the argument goes that by beginning early, before the path of the technological development becomes set, it will be possible to ensure that autonomous machines are able to self-regulate so as to serve the goals of IHL.¹⁸⁰

While this is indeed a possibility, it is only that, a possibility, and there are many obstacles, both technical and otherwise, that may prevent it from being the certainty that is assumed by those eager to incorporate AI into the military as soon as possible. The issue must be approached with caution. It is not guaranteed that the logic of AI will automatically, or even easily, be subjected to the logic of militaries, or IHL. The merging of these strands will inevitably result in a trade-off, a cost imposed, or a sacrifice made. In the case of IHL, the trade-off is between accepting the notion of violence in war, and then seeking to put a cost on the unrestrained use of violence. When it comes to AI, the value alignment trade-off should be the focal point from which minimum standards can be established, which will then be articulated generally enough that they can take account of future advances in the technology. While the problems of cognitive explanation and value alignment remain unsolved, the trade-off would be largely in favour of strong human oversight over the optimal performance of the systems. Should these various challenges be solved in the future, the trade-off would be more favourable towards reduced human oversight. The metric by which the level of oversight required would be measured are the principles of awareness and control underlying the doctrine of superior responsibility. How much awareness and control does the system allow a commander to maintain? In the way that a commander is required to train the troops in the knowledge of IHL during peacetime, so might he or she be responsible for ensuring that any AI decision-making systems are likewise trained and updated in respect of changing tactical situations, military doctrines, rules of engagement and international laws.

As for the responsibilities during MDMP, a commander may be required to assess whether or not a particular system is able to provide a satisfactory explanation of its own cognitive processes, such as how it arrived at a particular conclusion regarding a proportionality assessment. He or she will also have to be confident that the parameters of the systems’ architecture include IHL, and that it

¹⁷⁸ DARPA, *XAI* (n. 169)

¹⁷⁹ Conn, A., *How do we align Artificial Intelligence with Human Values*, (2017) FLI

¹⁸⁰ For example, Anderson and Waxman (n. 20) pp. 9, 27

can conform to them while still optimising for a successful operation. This will require of commanders a level of technical expertise that some may rail against. However military commanders are not legal experts and yet are still required to have knowledge of the law commensurate with their position and to involve legal advisors to supplement their knowledge. Given the transformational effect that AI will have on warfare in the future, a similar requirement may be established in this regard. Thus military commanders will have knowledge of AI that is commensurate with their function, and expert advisors to fill in any gaps. One can imagine a tripartite arrangement whereby an expert in the AI decision-making system, be it the programmer, engineer or other expert, advises the commander, together with the legal advisor, on the implications of the system's output. This is not a simple solution, given the general problem with machine learning algorithms moving beyond the understanding of their architects, but is worthy of further exploration in the future. Guidance on their role, similar to that issued by the ICRC for legal advisors, may be incorporated into the general arsenal of rules applicable to superiors, as part of the principle of responsible command.

An emphasis on minimum standards may not always be immediately satisfying for most lawyers and activists that are used to robust accountability mechanisms to ensure respect for the law. While this is a fair critique, it must be seen in light of the international system of humanitarian law and human rights as a whole. The nature of these bodies of law differs from domestic systems in that they depend significantly on spontaneous compliance and domestic enforcement, with some mechanisms of accountability for the most serious violations. As a result of the shortcomings of such a system, it relies on preventative measures to a much greater degree than domestic systems. The doctrine of superior responsibility is no different in that it seeks to prevent violations on the one hand through respect for IHL during the planning process, while holding commanders criminally accountable for failing to prevent and punish serious violations committed by their subordinates. The difference between superior responsibility and the broader IHL system is that this former function of superior responsibility has been significantly underdeveloped in comparison to preventative measures of IHL. The Geneva Conventions, for example, impose a duty on states to disseminate IHL to the relevant parties, and to adequately train its armed forces.¹⁸¹ There are non-proliferation treaties in respect of certain conventional weapons, requirements for the adoption of domestic legislation that criminalises violations of IHL and the protection of protected emblems and there is the emphasis that is placed on using legal advisors throughout the planning process. Kolb writes that *"it is hardly possible to over-estimate the practical importance of preventative measures. These may not be as visible as suppressive one, but have a much greater impact for the proper working of IHL than the latter, which by definition are always 'belated.'"*¹⁸² The rise of AI assisted decision-making brings this into the forefront in respect of the doctrine of superior responsibility which, with some work, is well placed to ensure that use of such systems does not negatively impact the implementation of IHL norms during the planning, and consequently the execution of warfare in its routine functioning.

5.3. Individual liability of Superiors

This thesis does not suggest that an increased emphasis on minimum standards applicable to superiors should replace the existing mechanisms by which individuals are held criminally responsible. What effect, then, will AI decision-making tools have on the individual liability of superiors, whether indirectly for the conduct of their subordinates, or directly for their own conduct under international criminal law?

Although neither of these bodies of law is explicitly limited to high-ranking officials, the international tribunals and prosecutors have trained their focus ever more increasingly on this category of persons. One reason stems from the view that persons placed higher in the command hierarchy are generally more responsible for the kind of large-scale atrocities that occur during war. Another reason is the expectation that by holding those higher up the chain of command responsible

¹⁸¹ Geneva Conventions I-IV, Articles 47/48/127/144

¹⁸² Kolb (n. 8) p. 190

for the control they maintain over their troops, a culture of accountability will filter down and so prevent persons lower in the hierarchy from committing crimes.¹⁸³ In order to understand how the liability of individuals is affected, it is useful to think about the different scenarios under which they can be prosecuted. The first is for the intentional, systemic commission of war crimes¹⁸⁴ that usually emanate from the highest level of military and political leaders, such as the crime of genocide. The other is for war crimes committed by subordinates, either pursuant of the commander's orders or which the superior failed to prevent or to adequately punish.

Each scenario engages the operation of either direct or indirect criminal responsibility, or both, and it is not envisaged that the use of AI decision-making tools will have a significant impact on the ability of international law to hold individuals to account in these cases of extreme violations of IHL and international criminal law in any of the above scenarios. The decision to intentionally engage in wide-spread, systemic war crimes are usually not made by low ranking tactical or even operational commanders but by the higher-level military and political leaders. The ability for AI systems to replace decision-making at these levels is a long way off which means that individual liability will continue to operate as it does now. The question of who and how individuals will be held accountable for the misuse of if AI systems that result in serious violations of IHL and war crimes will continue to be addressed under international criminal law. The accountability gap that is a major concern in respect of LAWS does not occur here as there is always a human to translate the plan into orders.

However, in respect of operational decision-making, this particular application of AI reveals the shortcomings of the concept of humans-on-the-loop and gives some insight into what constitutes meaningful human control, both of which are central features of the debate about LAWS. The use of decision-making systems will necessarily involve humans for the foreseeable future given that they do not have the capabilities to carry out operations directly. Tactical and operational commanders will, at the very least, maintain control over ordering, executing and monitoring the implementation of the plans recommended by these systems and thus there is always, in principle, an individual to hold responsible for the commission of unlawful acts during warfare. The challenge regarding decision-making tools, however, arises not in who to hold accountable but rather how to establish that they had the requisite *mens rea*, if they merely executed a plan, the implications of which they could not fully comprehend nor predict. This is similar to an existing problem within IHL, Clausewitz' fog of war, and the problem of friction. The fog of war, we know, is a result of imperfect intelligence regarding the conditions of the battlefield. Friction refers to the difficulty of maintaining control over how the execution of operations plays out in reality, given that everything from weather, to difficulties in supply and logistics, and to the reactions of enemy forces are out of the commander's control. In light of these two inherent characteristics of the battlefield, it becomes problematic for a prosecutor to establish that a commander had the intention to violate the laws of war in a given situation, long after the fact. These difficulties will most likely be amplified by the fog of AI presented by the black box problem.

Until now, lawmakers have sought to address this complication through the strict standards of *mens rea* applicable to direct individual criminal liability, and indirect liability under the superior responsibility, discussed in Section II. The underlying rationale for these bodies of law is to disallow impunity of those leaders who avoid responsibility for mass atrocities by having others carry out their criminal plans. The very purpose and duty of the higher-levels of command, besides running successful military campaigns, is to maintain control over the conduct of warfare and to ensure that any instructions given to subordinates comply with the limitations of IHL. It is implausible therefore that lack of insight into the cognitive processes of a system that has been intentionally used to conduct the MDMP would constitute a sufficient defence against prosecution for war crimes and grave breaches of IHL, particularly in light of the standard of *mens rea* under the ICC Statute. There is, however, a potential evidentiary problem in establishing the requisite *mens rea*, between the commander using a system and the programmer, which is flagged below.

¹⁸³ Mettraux (n. 11) p. 14-15

¹⁸⁴ The term 'war crimes' here also includes grave breaches of IHL

It is therefore envisaged that criminal liability should continue to be a vital mechanism in holding individuals responsible for war crimes and grave breaches of IHL as it does today. Individual criminal liability, both direct and indirect, are a hard won avenue for addressing impunity, which will not be undone by the simple fact that AI systems support the decision-making process. Knowledge of what constitutes war crimes and grave breaches is available and acting on the orders thereof will remain subject to criminal liability. Yet what the current system does not do is address all failures of responsible command with criminal liability. While this is true in principle, it is clear that there will be evidentiary problems in establishing the guilt of superiors, even in light of the lower standards of *mens rea* in respect of war crimes and grave breaches of IHL. It may be that use of AI decision-making systems results in fewer convictions in the future. Once again, this is an area for further exploration that goes beyond the scope of this thesis.

Criminal liability, or other forms of military discipline, that apply to lower ranking members of the armed forces will also continue to operate. Impunity in their case is protected against by the fact that merely following superior orders is not a valid defence against the commission of a war crime.

5.4. Individual Liability of AI experts

Finally, it is necessary to highlight the potential introduction of the AI programmer¹⁸⁵ into the line of accountability, which has in fact been a question in the discourse of AI regulation since the early 1990s. If AI decision-making systems are employed on the back of their capability to into account considerations of IHL, then common sense tells us that the person or persons who are responsible for creating and training the system are at least partly responsible for ensuring that it does so, and hence liable when it fails. There has been a range of proposals regarding how to extend liability to an AI programmer, both civil and criminal, given that there would be no unlawful act by him or her directly. One suggestion relies on the model of natural-probable-consequence which would allow a for the criminal or civil liability of a programmer for the misbehaviour of an AI that was otherwise intended for positive purposes if that misbehaviour was foreseeable as a natural and probable consequence of that system.¹⁸⁶ While this may work in relatively stable environments in which ‘misbehaviour’ is easily identifiable, it would be challenging in the dynamic, uncertain context of warfare. Another suggestion involves criminal liability under the model of perpetrator-via-another, which addresses the situation where an offence is committed by a party that cannot form the necessary *mens rea* (as a result of a mental deficiency of some sort), but that was instructed by another to commit the offence.¹⁸⁷ In that way it is analogous to the doctrine of superior responsibility; both concern the crime of a party other than the perpetrator. This model is attractive as it allows for malice or negligence on the part of the programmer, but there would nevertheless be a significant evidentiary burden in respect of highly subjective assessments that are required to be programmed long in advance of the critical moment of decision-making. To make this worse, if the system uses deep learning, it will ‘correct’ itself potentially beyond the programmer’s original intent. In such circumstances it is, in fact, the human programmer and user that becomes the mentally deficient innocent party insofar as they are unable to understand the internal logic of the AI and are merely following its advice. This section is a brief introduction to a much larger issue that requires further exploration but illustrates the broader truth that thinking about IHL norms will no longer be the sole province of politicians, lawyers and activists.

¹⁸⁵ Note that the issue is further complicated by the fact that in all likelihood there would have to be a sharing of responsibility between numerous parties involved in designing, manufacturing, providing data, training and overseeing the AI, as well as the entity that employs each of these individuals. The reference to ‘programmer’ should be read to encompass all of these parties.

¹⁸⁶ Kingston, J., ‘Artificial Intelligence and Legal Liability’, in Bramer, M., and Petridis, M., (eds.), *Research and Development in Intelligent Systems XXXIII: Incorporation Applications and Innovations in Intelligent Systems*, (2016), Springer, p. 272

¹⁸⁷ *ibid*

SECTION VI – CONCLUDING REMARKS

On the trajectory of AI, the technology has reached the point where computer systems are intelligent enough to share (if not yet replace) many of tasks previously thought to be exclusively human activities. When it comes to military decision-making in the U.S. armed forces, this balance is still weighted heavily in favour of humans. Current operationalisation into the military has an inverse relationship to the military command structure, with the concentration decreasing up the chain towards the operational and strategic levels, and yet there are a number of ways that the use of tactical AI influences the higher levels of decision-making, both indirectly, and directly. In the meantime, the technology continues to advance in its ability to conduct increasingly sophisticated decision-making and planning. Although presently confined, for the most part, to civilian applications, the technology is at least theoretically transferrable to the military domain because of its foundational nature. In light of the significant advantages it will bring to this particular area of warfare, it is certain that IHL will soon face a new paradigm of military decision-making whereby intelligent systems take over more and more of the MDMP. This will result in a clash of three fundamentally different projects - military, IHL and AI - which will inevitably require concessions and points of cohesion. The international law community must, therefore, engage in an intellectual exercise to apprehend where the three will intersect and to ensure that they do so in a way that does not compromise the fundamental principles of IHL.

Military normativity and principle of humanity have achieved this compromise through the principles of military necessity and responsible command. This latter principle is manifested by the doctrine of superior responsibility and, in the U.S. (and other nations) by the inclusion of legal advisors into the process of MDMP. AI adds another layer to this compromise. It has the potential to disrupt the division between strategic, operational and tactical decision-making, and consequently the legal structures that have been built around that model. Moreover, it is not yet clear whether AI systems will be able, or willing, to take account of IHL and, even if they are, many of the legal assessments that take place during MDMP are highly subjective exercises requiring experience and talent. Finally, the inscrutability of AI processes will render it difficult, if not impossible, for human superiors to ensure the implementation of the full spectrum of IHL norms during warfare, and to hold individuals accountable for serious violations that are a consequence of executing on its automated plans.

Because there is no treaty or provision of IHL that explicitly governs the conduct of military decision-making, nor any rules regulating the use of technology in the MDMP, achieving a ban on AI decision-making systems altogether is highly unlikely. It becomes necessary, then, to look to other frameworks of IHL to ensure that they are used responsibly. The mechanisms of individual liability in international criminal law only address specific, serious violations of IHL and fall short of guaranteeing compliance with the full spectrum of laws applicable during warfare. The focus of this thesis, therefore, has been on the broader duty of superiors to ensure that armed conflict is carried out in a legally compliant manner, as a consequence of their critical position at the threshold of warfare, and the control they exercise over subordinates. If anything, modern warfare has made this role even more significant as low-level superiors take ever more strategically important decisions. It is imperative, therefore, to establish a workable standard of principles that can adapt to and encompass future applications of AI, which has been the role of the general principles of IHL since its inception and must be so again in light of rapidly developing technologies.

Currently, the doctrine of superior responsibility is the body of rules and minimum standards that endeavours to realise the incorporation of IHL into the process of MDMP. The principle of responsible command underlying the doctrine, and the duties of awareness and control, can form the basis of minimum standards aimed at ensuring that AI is used responsibly during the prosecution of warfare. However, the doctrine will require amendment, augmentation, and supplementation, to take account of the use of AI in decision-making processes. This will involve a two-tiered approach

of expanding on and solidifying the preventative aspect of superior responsibility on the one hand and adapting the evidentiary requirements on the other. A significant component of the first tier will be the duty on superiors to ensure that they are adequately informed by experts on the capabilities of the systems, primarily during peacetime, and to assist him or her to understand the nuances of each output during armed conflict. Failure to do so would then constitute a breach of the duty to retain awareness and control over the conduct of operations, and of the general principle of responsible command. The ICRC can issue guidelines on the requirement for, and role of AI experts in the MDMP, as has been done in the case of legal advisors. The role of the legal advisor would also be updated to reflect the new paradigm. The second tier will depend largely on adapting the evidentiary proof required to establish *mens rea* to take account of situations whereby the output of AI systems result in serious violations of IHL and war crimes. So long as humans retain responsibility for translating automated plans into action through the issuing of orders and by maintaining control over execution thereof, ignorance following from the opaque cognitive process of AI systems can never be a defence against responsibility for the violation of the norms and standards of IHL. A dictate in this regard would be comparable to the rule that following superior orders is not a valid defence against the commission of a war crime under international law.

The upshot of using AI in the decision-making process is that an additional shroud of uncertainty is added to the existing fog and friction of war which will have significant implications for IHL. The cost that is imposed on military force by IHL cannot be allowed to be subsumed by the logic of AI. Instead, AI must be developed and used so as to conform to existing legal frameworks, in this case, the doctrine of superior responsibility. The unique challenge of IHL in the context of AI is to establish these frameworks in anticipation of its operationalisation and not merely in response to it, which is the rationale behind framing the issue as a question of responsible command and thus relying on the flexible minimum standards of the doctrine of superior responsibility.

BIBLIOGRPAHY

Books

- G. Best, *War and Law since 1945* (Oxford: Clarendon Press, 1994)
- R. Kolb, *Advanced Introduction to Humanitarian Law* (Cheltenham: Edward Elgar Publishing Limited, 2014)
- G. Mettraux, *The Law of Command Responsibility*, Oxford University Press, Oxford, 2009
- D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, (Leiden: Martinus Nijhoff Publishers, 2013)
- M. Tse-Tung, *Selected Military Writings of Mao Tse-Tung*, (Peking, PRC Foreign Languages Press, 1963)
- C. Von Clausewitz, *On War*, Edited and Translated by Michael Howard and Peter Paret, (Princeton: Princeton University Press, 1989)
- G. Werle, *Principles of International Criminal Law*, 3rd ed., (Oxford: Oxford University Press, 2014)

Chapters in Books

- K. Ambos, 'Superior Responsibility' in A. Cassese I. (eds.), *The Rome Statute of the International Criminal Court: A Commentary*, 1 (Oxford: Oxford University Press, 2002), 823
- L. Doswald-Beck, 'Some thoughts on the Computer Network Attack and the International Law of Armed Conflict', in M. Schmidt and B. O'Donnell (eds), *Computer Network Attack and International Law, Law Studies 76*, (Newport: Naval War College, 2002), 163
- J. Kingston, 'Artificial Intelligence and Legal Liability', in M. Bramer and M. Petridis, *Research and Development in Intelligent Systems XXXIII: Incorporation Applications and Innovations in Intelligent Systems*, (Springer, 2016) 269
- M. Kouba (ed.), *Operational Law Handbook* 17th ed., (Virginia: The Judge Advocate General's Legal Centre and School, 2017)
- D. Saxon, "A Human Touch: Autonomous Weapons, Directive 3000.09, and the "Appropriate Levels of Human Judgment over the Use of Force", in N. Bhuta *et al.* (eds.) *Autonomous Weapon Systems: Law, Ethics, Policy*, (United Kingdom: Cambridge University Press, 2016) 185

E-Books

- I. Goodfellow, *et al.*, *Deep Learning*, (MIT Press, 2016) <<http://www.deeplearningbook.org>>
- S. Marsland, *Machine Learning an Algorithmic Perspective*, 2nd ed., (CRC Press, 2015) <<https://www.crcpress.com/Machine-Learning-An-Algorithmic-Perspective-Second-Edition/Marsland/p/book/9781466583283>>
- R. Wallace, *Carl von Clausewitz, the Fog-of-War, and the AI Revolution: The Real World Is Not A Game Of Go*, (SpringerBriefs, 2018) <<https://link.springer.com/content/pdf/10.1007%2F978-3-319-74633-3.pdf>>

Cases

- Case concerning the Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion*, ICJ Reports, 1996, para 78
- Prosecutor v Boškoski and Tarčulovski*, Case No. IT-04-82-A19, Judgment 19 May 2010, AC Decision
- Prosecutor v Hadžihasanović et al*, Case No. IT-01-47-A, Judgment 22 April 2008, AC Decision

Articles

- R. Acheson (ed.), *Losing Control: The Challenge of Autonomous Weapons for Laws, Ethics, and Humanity*, *CCW Report on the CCW Group of Governmental Exports on Lethal Autonomous Weapons Systems*, (2017) 5(3) Women's International League for Peace and Freedom, 13
- K. Anderson, M. Waxman, *Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Laws of War Can*, (2013) Stanford University, The Hoover Institution (Jean Perkins Task Force in National Security and Law Essay Series)
- R. Arkin, P. Ulam and A. Wagner, *Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception*, (2011) 100(3) Proceedings of the IEEE
- R. Arkin, *Warfighting Robots Could Reduce Civilian Casualties, So Calling for a Ban Now is Premature*, (2015) IEEE Spectrum,
- P. Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, (2012) 94 Int'l Rev. Red Cross, 687
- K. Ayoub, K. Payne, *Strategy in the Age of Artificial Intelligence*, (2016) 39(5-6) Journal of Strategic Studies, 793
- R. Betts, *Compromised Command*, (2001) 80(4) Foreign Affairs (Review of Wesley Clark's, *Waging Modern War: Bosnia, Kosovo, and the Future of Combat*)
- Thomas Burri, *International Law and Artificial Intelligence*, (2017) German Yearbook of International Law 2017/18 (Forthcoming)
- M. Cummings, *Man versus Machine or Man + Machine?*, (2014) 29(5) IEEE Intelligent Systems, 62
- M. Cummings, *Artificial Intelligence and the Future of Warfare*, (2017) Chatham House (Forthcoming Chatham House report on artificial intelligence)
- J. Dunnaback, *Command Responsibility: A small-unit leader's perspective*, (2014) 108(4) Northwestern University Law Review, 1385
- L. Ground, A. Kott, R. Budd., *Coalition-based Planning of Military Operations: Adversarial Reasoning Algorithms in an Integrated Decision Aid*, (2015) arXiv: 1601.06069 Computing Research Repository
- A. Kott, R. Budd *et al.*, *Decision Aids for Adversarial Planning in Military Operations: Algorithms, Tools and Turing-test-like Experimental Validation*, (2016) arXiv:1601.06108, Cornell University Library and Applied Intelligence Journal
- D. Lewis, G. Blum, N. Modirzadeh, *War-Algorithm Accountability*, (2016), HLS Program on International Law and Armed Conflict, Research Briefing
- M. Lohr, S. Gallotta, *Legal Support in War: The Role of Military Lawyers*, (2003) 4(2) Chicago Journal of International Law, Article 14
- T. Meron, *The Humanization of Humanitarian Law*, (2000) 94(2) AJIL, 239
- V. Mnih *et al.*, *Human-level control through deep reinforcement learning*, (2015) 518, Nature, 529
- A. Ng *et al.*, "Autonomous Inverted Helicopter Flight via Reinforcement Learning", in M. Ang and O. Khatib (eds.), *Experimental Robotics IX: Springer Tracts in Advanced Robotics*, (2006) 21, Springer, 363
- R. Rasch, A. Kott, K. Forbus, *Incorporating AI into Military Decision Making: An Experiment*, (2003) 18(4) IEEE Computer Society July/August edition, 18
- H. Roff, *Advancing Human Security through Artificial Intelligence*, (2017) Chatham House (Forthcoming Chatham House report on artificial intelligence)

Letters, Memoranda, Presentations

- Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, (2015)
<<https://futureoflife.org/open-letter-autonomous-weapons#signatories>>
- An Open letter to the United Nations Convention on Certain Conventional Weapons* (2017)
<<https://futureoflife.org/autonomous-weapons-open-letter-2017/>>
- Col. C. Dunlap, Jr., *Law and Military Interventions: Preserving Humanitarian Values in 21st Century Conflicts*, Conference Paper for Humanitarian Challenges in Military Intervention

- Conference, Carr Center for Human Rights Policy, Kennedy School of Government, Harvard University, 2001
- A. Kott and P. Corpac, *COMPOEX Technology to Assist Leaders in Planning and Executing Campaigns in Complex Operational Environments*, 12th International Command and Control Research and Technology Symposium, 2007
- W. Parks, Office of The Judge Advocate General of the Army, *Weapons Review Programme of the United States*, presented at the Expert Meeting on Legal Reviews of Weapons and the SIRUS Project, Switzerland, 2001
- J. Scholz *et al.*, *Machine learning for adversarial agent microworlds*, MODSIM 2005 International Congress on Modeling and Simulation, Modeling and Simulation Society of Australia and New Zealand, 2005

Specialist Reports

- M. Brundage, S. Avin, *et al.*, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, Centre for the Study of Existential Risk, *et al.*, February 2018
- A. Conn, *How do we align Artificial Intelligence with Human Values*, (2017), Future of Life Institute, <<https://futureoflife.org/2017/02/03/align-artificial-intelligence-with-human-values/>>
- J. Ellman, *et al.*, *Assessing the Third Offset Strategy: A Report of the CSIS International Security Program*, (2017), Centre for Strategic and International Studies
- Group of Governmental Experts of the High Contracting Parties to CCW, *Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems* (2017) Geneva, (LAWS) CCW/GGE.1/2017/CRP.1
- C. Heyns, *Report of the Special Rapporteur on extrajudicial summary or arbitrary executions*, Adopted by the UN General Assembly, Human Rights Council, at its 23rd Session, 9 April 2013, 23(3). A/HRC/23/47
- Human Rights Watch and the International Human Rights Clinic at Harvard Law School, *Losing Humanity: The Case Against Killer Robots*, (2012), ISBN: 1-56432-964-X
- United States Government Accountability Office, *Technology Assessment: Artificial Intelligence Emerging Opportunities, Challenges, and Implications*, (2018) Report to the Committee on Science, Space, and Technology, House of Representatives, GAO-18-142SP.

ICRC commentaries, manuals and guidelines

- F. De Mulinen, *Handbook on the Law of War for Armed Forces*, (1987) ICRC, Geneva
- J. Henckaerts and L. Doswald-Beck, *Customary International Humanitarian Law*, Volumes I and II (Cambridge: Cambridge University Press, 2009)
- ICRC, *Legal Advisers in the Armed Forces*, (2003) Advisory Service on International Humanitarian Law
- ICRC, *Handbook on International Rules Governing Military Operations*, (2013) Geneva
- ICRC, *Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, (2006) 88(864) Geneva
- Y. Sandoz, C. Swinarski, B. Zimmermann (eds.), *Commentary on the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, (Geneva: Martinus Nijhoff Publishers, 1987)

United States military documents

- Department of the Army
- FM 100-5, Operations, (1993)
- FM 101-5, Staff Organization and Operations, (1997)
- Regulation 600-20, Military Discipline and Conduct, (2006)

Department of Defense

Directive 5100.77, DoD Law of War Program, (2006)

Directive 3000.09, Autonomy in Weapons Systems, (2012)

Joint Chiefs of Staff

Joint Publication 3-60, Joint Doctrine for Targeting A-6, (2007)