

JANAR PEKAREV

Military AI and autonomous
weapon systems: an interdisciplinary
exploration of ethical, legal,
and sociological aspects



JANAR PEKAREV

Military AI and autonomous weapon systems:
an interdisciplinary exploration of ethical,
legal, and sociological aspects



UNIVERSITY OF TARTU

Press

Institute of Social Studies, University of Tartu, Estonia

Dissertation accepted in fulfilment of the requirements for the degree of Doctor of Philosophy (in Sociology) on April 23, 2024, by the Council of the Institute of Social Studies, University of Tartu.

Supervisor: Veronika Kalmus, Professor
University of Tartu, Estonia

Opponent: René Moelker, Associate Professor of Sociology
Netherlands Defense Academy

Commencement: May 29, 2024, Senate Hall of the University of Tartu

The publication of this dissertation is granted by the Institute of Social Studies, University of Tartu.

ISSN 1736-0307 (print)
ISBN 978-9916-27-523-8 (print)
ISSN 2806-2590 (pdf)
ISBN 978-9916-27-524-5 (print)

Copyright: Janar Pekarev, 2024

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL STUDIES.....	7
ACKNOWLEDGEMENTS	8
INTRODUCTION.....	9
1. THEORETICAL FRAMEWORK	13
1.1 Breaking down AI-driven warfare: the anatomy of autonomous systems.....	14
1.1.1 The definition – the first caveat.....	14
1.1.2 The binary problem – the second caveat.....	15
1.1.3 Autonomous use of force – the third caveat.....	16
1.1.4 Insight into AI weaponisation	19
1.2 The conceptual basis of the three pillars – ethical, legal and sociological aspects.....	20
1.2.1 Ethical aspects – the first pillar	21
1.2.2 Legal aspects – the second pillar.....	22
1.2.3 Sociological aspects – the third pillar	23
2. METHODOLOGICAL CONSIDERATIONS.....	25
2.1 Data and Methods (Study I – Review).....	25
2.2 Data and Methods (Study II – Focus Groups)	26
2.2.1 Ethical Matters	27
2.2.2 Sample, Data Collection and Analysis.....	27
2.3 Data and Methods (Study III – Interviews)	29
2.3.1 Sample, Data Collection and Analysis.....	29
2.4 Data Visualisation in the Findings.....	31
2.4.1 Similarity Analysis of Qualitative Data	31
3. FINDINGS	33
3.1 A Compendium of Arguments (Study I).....	34
3.2 A focus-group perspective (Study II).....	35
3.2.1 Distinguishing Human and Machine Decision-Making.....	36
3.2.2 Assigning Responsibility Involving Autonomous Weapons.....	38
3.2.3 Arguments for and Against the Development of Autonomous Weapons	39
3.2.4 Discomfiture over Weaponised AI.....	40
3.3 Insights from Military Experts (Study III).....	40
3.3.1 Military applications of warfare in autonomous mode	41
3.3.2 The Unknown in Unmanned Warfare	44
3.4 A Convergence of Thoughts: Citizen and Expert Insights on Autonomous Weaponry	46
3.4.1 Unity in Uncertainty: Bridging Perspectives on Unmanned Warfare.....	47

4. DISCUSSION	49
4.1 Decisions of AI-driven Autonomous Systems in the Military Domain.....	50
4.1.1 ‘Knowledge’ – Awareness that a Consequence will Occur	51
4.1.2 The Pivotal Moment: Moral Judgements in Crucial Scenarios.	53
4.1.3 A ‘Black Box’ – Input and Output of the Mind and Algorithm	56
4.1.4 A Human-Machine Teaming.....	58
CONCLUSIONS	62
REFERENCES.....	68
SUMMARY IN ESTONIAN	76
PUBLICATIONS	83
CURRICULUM VITAE	142
ELULOOKIRJELDUS.....	144

LIST OF ORIGINAL STUDIES

This dissertation is based on the following original publications, which will be referred to in the dissertation by their respective Roman numerals.

- Study I:** Pekarev, J (2021). Working towards ethical autonomous weapon systems: a compendium of arguments. *Sõjateadlane (Estonian Journal of Military Studies)*, 18, 184–202.
- Study II:** Pekarev, J (2023). Dilemmas of representation in a study of autonomous weapon systems. A focus group perspective. *Portuguese Journal of Military Sciences*, 221–240.
- Study III:** Pekarev, J (2023). Attitudes of military personnel towards unmanned ground vehicles (UGV): a study of in-depth interview. Springer: *Discover Artificial Intelligence*.

Author's Contribution

As the author of this dissertation, I have contributed to these studies as follows:

- Study I:** I am the sole author of the study.
- Study II:** I am the sole author of the study. Data gathering was conducted collaboratively with a research group from the Estonian Military Academy.
- Study III:** I am the sole author of the study. Data gathering was conducted collaboratively with a research group from the Estonian Military Academy.

ACKNOWLEDGEMENTS

It is with immense gratitude that I acknowledge the support and assistance of those who made the journey of this dissertation not only possible but a meaningful experience. This thesis is a testament to my efforts and the invaluable contributions of those who have guided, inspired, and stood by me throughout this academic endeavour.

First and foremost, I extend my most profound appreciation to my supervisor, Professor Veronika Kalmus, whose professionalism and patience have been integral to the completion of this dissertation. Her exceptional knowledge, professional rigour, and insightful feedback have enriched this research and significantly enhanced my understanding of the importance of scientific communication. Researching alongside Professor Kalmus has been a privilege. Her steadfast support and perceptive critiques have played a vital role in my evolution as a junior researcher in a manner crucial to the success of this thesis.

I would like to express my profound gratitude to Professor Kairi Kasearu and Tiia-Triin Truusa (PhD) for their invaluable support and guidance. Their encouragement inspired me to pursue this scientific journey and gave me the fortunate opportunity to be introduced to my supervisor. I am immensely thankful to Colonel Raul Järviste, former Head of the Department of Applied Research at the Estonian Military Academy. His influential role and mentorship have been crucial in steering my academic career path. The department's collective efforts have contributed to my professional growth and achievements, culminating in the development of my thesis topic, a subject I am excited and eager to share.

I extend my heartfelt thanks to my colleagues at the iMUGS project. The data and insights gathered from this project have been pivotal in analysing unmanned ground systems. I am particularly indebted to Professor Wolfgang Wagner, Kairi Talves, Eleri Lillemäe, Auli Viidalepp, Markus Otsus, Camilla Guldahl Cooper, and Cecilie Hellestveit. I am also profoundly grateful to the NATO SAS-160 members and team leader, Ms Linda Holtby, for their invaluable contributions to exploring emerging military technologies. Their efforts in shedding light on new perspectives have been instrumental and have led me to re-evaluate various aspects for inclusion in my thesis.

My family's unwavering support and constant encouragement, patience, and emotional backing have given me the strength and stability needed to navigate the rigorous demands of academic research. I am deeply grateful for their love and support, which have been invaluable in helping me reach this milestone in my career. In particular, I wish to devote this to my beloved Sisi. Your presence has been a beacon of light and inspiration throughout my doctorate studies. Your innocent questions about my progress and gentle nudges to keep me focused on work and reading have been a true motivation. Your curiosity and concern about technology, as I delved into the realm of military AI, are reflected in your thoughtful and imaginative drawings of robots, adding a unique and heart-warming perspective to my work. With all my love and gratitude, I dedicate this achievement to you, my little star.

INTRODUCTION

*This issue must be confronted
if there is any agreement as to what
machine ‘understanding’ might mean.
– Joseph Weizenbaum (1966)*

One of the pioneers and a critic of artificial intelligence (AI), Joseph Weizenbaum, postulated in the mid-1960s that the ultimately responsible human interpreter of ‘What the machine says’ is constantly faced with the need to make credibility judgements on how easy it is to create and maintain the illusion of understanding (Weizenbaum, 1966). He emphasised that while machines can simulate understanding, they lack genuine comprehension, empathy, and ethical reasoning. Half a century later, in turbulent times in computer science, making sense of machine interpretation has become more puzzling than ever. More specifically, the AI explosion is reminiscent of the internet in its early days, notably outgrowing its military roots. The people working on the internet barely imagined how revolutionary a network of interconnected computers would be. The consequences AI could have on the military provoke many concerns, and as long as artificial decisions on the use of force are not excluded, it is very much relevant to consider hypothetical unmanned warfare. The war involving military AI, which appears to have already begun, is manifest in Ukraine and Gaza, as we have come to recognise by 2024.

In this thesis, the term military AI is broadly defined to encompass a range of applications within the military domain, specifically for exploring aspects of autonomous weapon systems (AWS) without delving into machine learning methods or individual technological models. I identify military AI as the critical enabler of intelligent functionalities in autonomous systems. The author’s position on these terms is crucial for navigating through the nuances of this dissertation; therefore, semanticising ‘autonomous’ and ‘unmanned’ serves as the key to demystifying the complexities inherent to AI in the military domain.

I will delve deeper into the subtleties of these terms at a later stage (see 1.1.3). To start, it is crucial to grasp that within this context ‘autonomy’ in weapons is akin to the functioning of a brain, providing the ability to make decisions and govern independently via algorithms. In contrast, ‘unmanned’ denotes more a physical platform without a human operator, which can be seen as the body that executes tasks, much like a human body operates under the guidance of its brain. Unmanned warfare has more broadly secured its position within the literature, which ultimately means that the use of force lies outside direct human attendance (Scharre, 2018).

Having said that, the tempo of recent AI adoption can be seen through its widespread integration across various industries, and the military is no exception (McFarland & Assaad, 2023). It could be argued that ignoring the capabilities of AI to alleviate the limitations of human cognitive performance would be irresponsible and unethical (Meerveld et al., 2023). Speculations that AI might surpass

human intelligence and initiate harmful scenarios carry particular gravity in the military domain, where the potential consequences are drastically magnified, and the risks of such advancements are viewed with heightened urgency (Dresp-Langley, 2023). Therefore, many voices suggest banning the development of so-called killer robots (Wareham, 2020) and alarmingly stress the need for reconsideration at the United Nations level (GGE on LAWS, 2023). Still, it is most likely unrealistic to stop scientific progress or impractical to enforce banning measures because of the magnitude of the global momentum driving AI research.

The utilisation of unmanned force, a vital aspect of multi-domain warfare, is examined through the lens of military sociology, a perspective essential for understanding these developments within a broader sociological context. Employing an exploratory approach is particularly suited to the early stages of studying the emerging field of military AI. War's historical impact on social structures across cultures and epochs further underscores the relevance of a sociological approach (Kleykamp & Yastrzemsky, 2017). The methodologies in military sociology are as varied as the field itself, including theoretical, philosophical, and practical approaches to address its broad spectrum of research problems and questions (Shields, 2020). My work, influenced by my connections to the armed forces, adopts an applied perspective in line with military sociology and integrates interdisciplinary considerations to contribute to military science and technology.

That being so, research on the utilisation of AI in the military domain is critical to raise awareness and expand the discussion. Moreover, monitoring public attitudes towards disruptive technologies to understand public opinion has become increasingly important in informing decisions about the development, regulation, and use of AI (Maslej et al., 2023). Therefore, this dissertation intends to offer a compact insight into an AI weaponisation debate considering significant ethical, legal, and sociological aspects resulting from an analysis of the military perspective. The aim is to present novel knowledge regarding the representations of military AI discussed by ordinary citizens and military experts.

With that in mind, I decided to tackle military AI issues, starting with a systematic literature review to grasp the ideas, set the scene and form the basis for the subsequent articles. Due to the lack of relevant data at that time, it was essential to systematise the current research at the start. The thesis comprises three distinct papers, each originating from individual research. The first paper (**Study I**) reveals empirical findings and scholarly opinions in academic journals on the problematic aspects of AI in the military domain. The findings of Study I indicated that the primary worries revolve around entrusting life-and-death decisions to an artificial entity and the subsequent legal challenges. At the same time, the literature review highlighted aspects that should be examined from a sociological point of view. The subsequent studies gathered data on how different segments of society view the potential benefits or threats from military AI.

An interdisciplinary approach is essential to developing a foundational framework for understanding and researching military AI, integrating ethical and philosophical considerations, sociological and psychological insights, legal and policy perspectives, and military operational analyses. Ethical explorations examine the

moral implications of unmanned warfare, while sociological assessments, gained through focus groups with ordinary citizens, shed light on sociological attitudes and impacts. Legal perspectives are crucial for understanding the regulatory landscape, and technology in real-world scenarios can delve into AI's capabilities and limitations. Operational considerations informed by military expert interviews evaluate the broader implications of autonomous systems. These diverse viewpoints and the synthesis of expertise aims to create a comprehensive and nuanced understanding of military AI's multifaceted implications. Adopting a multi-pronged approach that integrates various disciplines can provide a holistic insight into human attitudes and beliefs surrounding military AI. It is essential to remain open-minded and adaptive as this field rapidly evolves with shifting societal perspectives.

It could be argued that failure to recognise public opinion about this topic risks dramatically failing to maintain the moral centre of gravity in military affairs (Boylan, 2015). For this reason, the second paper (**Study II**), mainly inspired by social representation theory, focuses on understanding the opinions and concerns of how ordinary citizens think and talk about weaponised AI. The questions of the focus-group discussion addressed the implications of using AI for military purposes and encouraged participants to imagine the consequences resulting from that exploitation.

To make it possible for all the main parties to be heard, especially from an end-user standpoint, the third paper (**Study III**) explores how military experts envision the role of AI in the armed forces. In order to gain new ideas and perspectives, in-depth expert interviews provide a greater understanding of more significant insights into the reasons underlying the beliefs, thoughts, and attitudes of military experts. Military ethics played a central role in approaching the exploration of the interview transcripts to identify the contextual meanings of what military personnel, as a primary source, have to say about the AI-driven use of force.

I acknowledge that there has been an influx of papers researching particular AI applications in the field of security and defence. I am also guided by the notion that warfare has evolved into an intricate societal phenomenon, while AI technologies facilitate the immediate execution of military operations, leading to diverse aspects that should be examined through the interplay of multiple perspectives. The ambition of this dissertation is to complement recent studies with an interdisciplinary approach focusing on ethics, law and sociological aspects resulting from the original papers (**Studies I–III**). The cover article presents a novel perspective to make determined efforts to deal with the most significant issues arising from previous research and poses research questions (RQs), centring around the main topics in the literature as follows:

RQ1: What are the main aspects associated with introducing autonomous weapons, as identified in the systematic review? (**Study I**)

a. What is the argumentation underlying these aspects? (**Study I**)

RQ2: What constitutes autonomous weapons systems, and how is weapon autonomy conceptualised? **(Study I–III)**

- a. What are the characteristics of decision-making made by machines and humans, mainly concerning matters of life or death, mirrored in the views of the study participants? **(Study II–III)**
- b. Who, from the perspective of study participants, would be potentially responsible if war crimes are committed due to using autonomous weapons, and why is there a possibility for an accountability gap in such cases? **(Study II–III)**
- c. Why do participants consider the development of autonomous weapons ethically sound or morally unacceptable, and what are the argumentative reasons for moving forward with authorisation of the development or bringing pressure to bear on a ban? **(Study II–III)**
- d. What dilemmas arise in the era of weaponised AI among ordinary citizens? **(Study II)**

RQ3: What, according to expert assessments, is the projected impact of autonomous systems and the conduct of future military engagements? **(Study III)**

- a. What causes uncertainty about the algorithmic rationale in autonomous systems? **(Study III)**
- b. What is the unknown the military is expected to mitigate in unmanned warfare? **(Study III)**

RQ4: What are the converging opinions on autonomous weapons based on the insights of ordinary citizens and experts? **(Study II–III)**

1. THEORETICAL FRAMEWORK

Warfare has, until recently, been a subject involving human beings. People have become used to the idea that the soldier plays a central role in the armed forces. The uninhabited military seems unthinkable, at least for the foreseeable future – more of a thought experiment than an imminent reality. However, as technology advances and the nature of warfare evolves, this balance may shift, making what seems unthinkable today a topic of serious debate tomorrow. Lately, unmanned roles in the military are increasingly emerging alongside humans (Chen, 2019).

This dissertation is about autonomous systems for military use, which, by design, can be understood as systems primarily meant to function without human presence. That does not mean humans necessarily cannot intervene at some point, but the state-of-the-art is oriented at weapon systems being able to execute tasks independently. In that vein, proactively diving into relevant studies for this dissertation, the core issue has been using military force when personnel are absent.

It is crucial to note that the participant perspectives in this thesis (**Study II–III**) were collected in 2021, predating the Russian invasion of Ukraine. However, by 2024, the recognition of AI-powered technology (i.e., drones) as an integral component of conventional warfare has been firmly established in people’s minds. A manifestation of the importance of military AI is one of NATO’s key challenges in effectively utilising military instruments of power (MIOp) in compressed decision-making cycles due to integrating AI and autonomous weapons, aiming for strategic impacts (SFA23, 2023).

While acknowledging that a mono-disciplinary approach might lack thoroughness and detail, I utilised an interdisciplinary scale to broaden the discussion. As a result, I adopted a three-pillar interdisciplinary approach, which is structured around key aspects identified in the literature (**Study I**). This approach interlinks disciplines while examining AI-enabled autonomous systems through the lens of military sociology.

This chapter, therefore, is structured into sections introducing the complexities of AI in warfare and outlining the profound implications of the framework of this thesis. Each sub-chapter delves into the nuances of autonomous systems, starting by addressing the definitions that shape my understanding of military AI capabilities and behaviours. The following paragraphs encompass the system’s physical components and software, including a discussion of the philosophical and technical aspects of autonomy and AI’s relation to lethal force. The second half of the chapter concludes by emphasising the importance of an interdisciplinary approach in understanding the complex interplay of ethical, legal, and sociological considerations in the context of AI-driven warfare.

1.1 Breaking down AI-driven warfare: the anatomy of autonomous systems

The integration of AI into warfare is not just a technological shift; it has profound implications for ethics, law, geopolitics, and the very nature of war. AI represents a significant transformation in the landscape of modern warfare and defence strategies. These systems, devoid of onboard human presence, combine the latest advancements in robotics, AI, and communications to execute a wide range of tasks, from surveillance and situational awareness to active combat roles. The anatomy of systems encompasses their physical components and the intricate software, meaningful control mechanisms, and the moral considerations underpinning their lawful operation. Understanding their composition, capabilities and implications becomes paramount as the military worldwide increasingly integrates autonomous systems into its defence arsenal. In conceding the abstractness of the subject, it is relevant to introduce three caveats that facilitate framing and exploring the subject matter for this dissertation.

1.1.1 The definition – the first caveat

To the best of my knowledge, a comprehensive and concrete definition of military AI is notably absent, making the existing terminology largely debatable and insufficient for fully understanding the phenomenon of the unmanned use of force. One of the first definitions of autonomy in weapons states that the system can select and engage targets without further intervention by an operator (Directive 3000.09, *Autonomy in Weapon Systems*, 2012). Nevertheless, this definition explicitly highlights the freedom of manoeuvring kinetic weapons. I will elaborate on the autonomous functions of military systems in a broader and more detailed context later (see 1.1.3).

Although I next explain what could embody autonomous systems, these descriptions must still be treated cautiously. The meaning and scope of words and concepts change constantly and probably differ from current understanding due to shifts in scientific advancements, which will improve essential nuances or subtleties associated with the manifestations of AI in arms. Therefore, the characterisation of autonomous systems is, in my mind, open to interpretation because it is deeply rooted in one's beliefs about AI's inherent capabilities and potentialities. For instance, if someone is convinced that AI can only execute tasks it has been explicitly trained for, they will likely perceive and characterise the system as rigid and deterministic. Conversely, if they believe that AI can adapt, learn from context, and modify its actions dynamically, they will see the system as more unpredictable.

In essence, the way that individuals make sense of and interpret the same system differently, in my view, highlights that the influence of personal biases and the depth of one's knowledge become apparent when considering attributions of human agency to AI. These subjective perspectives shape the way we predict and interpret the capabilities and behaviours of AI-driven systems. Recognising and

integrating diverse viewpoints is crucial, as they can profoundly impact the conclusions we reach about the system's potential outcomes.

A nuanced approach that considers these varied perceptions ensures that our expectations and assessments of AI are not just speculative but are grounded in an individual understanding. For instance, particular conclusions come down to whether creating machines with human-like cognitive capacities is possible; if the system qualifies as any other conventional weapon in someone's imagination, then AI-driven ethical concerns may be irrelevant (Coeckelbergh, 2020).

1.1.2 The binary problem – the second caveat

When considering the human mind and body analogy for AI, it is essential to differentiate between the 'mind' (software) and the 'body' (hardware). In many instances, AI exists predominantly as software, running on servers or in the cloud, without direct interaction with the physical world. When AI software is integrated into physical devices or machinery, it takes on a form reminiscent of the human mind-body connection. A self-driving car, for example, is a fusion of AI (the 'mind' making decisions) and the car hardware (the 'body' executing those decisions). The sensors act like the senses, collecting data from the environment, while the AI processes this data to navigate and drive, much like a human mind would process sensory input to manoeuvre in traffic. If we extend the analogy, AI learning from vast amounts of data could be seen as an accelerated form of how human minds gather knowledge over time. Meanwhile, the hardware's limitations and capabilities can shape AI performance, much as the human body's capabilities and constraints shape our actions.

Even if we focus on AI as software, it always requires some form of hardware to run, whether a server, a computer, or specialised machinery. However, in cases like cloud-based AI services, the hardware is often abstracted away from the user, making the 'body' less conspicuous. Autonomous systems are, in every way, a combination of hardware (aerial, ground, naval platforms) and software (machine learning to process data, large language and generative models). In this respect, the system can be a physical apparatus on the one hand and a theoretical decision-making instrument on the other. This combination allows AI to interact with and influence the real world in ways reminiscent of the human mind-body connection.

Most of the research has been conducted on drones, also known as unmanned aerial vehicles (Mozaffari et al., 2019; Zeng et al., 2016), and less on unmanned maritime vessels (Schmitt & Goddard, 2016) and much less on vehicles intended for terrain (Balch & Arkin, 1998). Furthermore, the concept of autonomous systems has also been associated with ground vehicles that navigate complex off-road environments (Young et al., 2017). This thesis focuses on an example of terrestrial platforms, specifically unmanned ground systems (UGS) or unmanned ground vehicles (UGV), to study autonomous systems.

The decision to emphasise terrestrial platforms was influenced by the design study from the iMUGS project (see 2.2.2), which offered a tangible example of autonomous weapons (a video of a UGV) aligned with the project's research

objectives. However, this does not imply that videos of aerial or maritime models, or even systems in a cyber domain, would have been unsuitable; they could also have provided insights into the study of autonomous systems. Due to the project's focus on a ground vehicle scenario, detailed observations of other domains were not conducted.

A video of a UGV operating in a forest area was used as the central representation for the participants (**Study II**) to demonstrate the combination of autonomy, chiefly to demonstrate the deep integration of hardware (the physical 'body') and software (algorithmic 'brain'). Regarding the software, algorithms are an integral part of any physical platform. For that reason, autonomous weapons can be analogised to AI-driven military platforms in relation to traditional human-operated units, such as a battle tank and its crew. This comparison serves to highlight the parallels between the cognitive and physical functions of humans (mind and body) and the computational and mechanical operations of AWS (software and hardware).

However, I remain open to the idea that algorithms could be classified solely as autonomous systems without a tangible physical component. The classification depends on the software's functionality and the intricacy of its decision-making processes. Systems that merely adhere to pre-established rules or decision trees and rigidly follow pre-defined instructions without learning from new data are not considered here. Actual autonomous systems should possess the ability to refine their decision-making based on new data or feedback. They are characterised by their capacity to manage intricate decisions, particularly when they must assess and prioritise multiple real-time variables. In essence, these algorithms excel in data analysis, pinpointing patterns, anomalies, trends, and threats, thereby directing machine actions or aiding human decision-makers (Scharre, 2023).

1.1.3 Autonomous use of force – the third caveat

The concept of weapon autonomy is grounded in extensive scholarly research, showing that it is well-supported by academic literature (Kanellakis & Nikolakopoulos, 2017; Macrina et al., 2020; Sharkey, 2019). The Department of Defence directive, referred to earlier (see 1.1.1), balances the benefits of leveraging advanced technologies with the necessity of maintaining control, responsibility, and ethical considerations in their application. However, there are inherent challenges and ambiguities in operationalising the autonomy function. The directive provides a framework, but real-world scenarios can make interpretations and applications way more complex. This means that the issue of autonomy in weapon systems remains a subject of debate and ongoing exploration.

The concept of meaningful human control has become known in the context of autonomous military applications, emphasising the crucial need for human oversight and ensuring that critical decisions involving lethal force remain under human judgement (Roff, 2013). Even so, it seems paradoxical that for as long as there is human control, to some extent, there cannot be full autonomy for machines (Bhuta et al., 2016).

It is worth mentioning that the discussion about machine autonomy in the philosophical sense of having free will and moral agency has been extensively debated and analysed in the literature (Cibralic & Mattingly, 2022; Gómez de Ágreda, 2020; Santoni de Sio & van den Hoven, 2018). However, I do not rule out that machine agency can be examined in specific contexts, but such a possibility does not significantly limit the present work. Moreover, this cover article tries to avoid direct anthropomorphism.

Autonomy chiefly points to the technical characteristics of machines and not the replication of human cognition. A machine's independence in executing complex tasks without human intervention refers to matching patterns in sensory input with pre-recorded data and the ability to select a course of action based on pre-defined objectives and selection criteria (Goussac et al., 2020). Otherwise stated, machine-independent reasoning could be seen as a sequence of cause-effect relations from capturing information to the execution of action (Dubber et al., 2020).

More remarkably, the concept of autonomy is most often associated with the use of lethal force (Asaro, 2012; Horowitz, 2019; Nyholm, 2018; Sparrow, 2016). The abstract idea of weapons autonomy has formed the widespread impression of AWS, also known in more morbid tones as *killer robots* (Rosert & Sauer, 2019; Simpson & Müller, 2016). In essence, the third caveat addressed in this thesis emerges from considering the autonomous capability of weapon systems.

This means that although the existence of the algorithmic use of lethal force is arguable, the dissertation was born from the idea that weapon systems will be eventually developed in an autonomous mode; that is, programmed and trained to operate independently. One might say it is inconceivable, but someday, such agents will act absolutely without human intervention and on their own behalf (Dubber et al., 2020). Be that as it may, to contest the future is necessarily to contest the present (Bhuta et al., 2016).

AWS often come with the adjective *lethal*, while the definitions of autonomous and unmanned vary inconsistently in the literature, causing significant confusion regarding any weapon system labelled autonomous (Del Monte, 2018). On that account, from what I have gathered, AWS can also be represented as machinery that perceives the surrounding environment, makes sense of it to some extent, identifies lawful targets, and is not only able to suggest the next move or make decisions on its own but also possesses the capabilities to apply lethal force beyond human supervision.

Despite efforts to explain this, I admit that the essence of autonomous weapons remains a contentious issue. Elaborating on the third caveat aims to clarify the rationale and highlight the autonomous functions as I perceive them. Consequently, some also distinguish between 'autonomous' and 'automatic' systems, which might exist in engineering (Lundberg & Johansson, 2021; Soltanzadeh, 2022). However, regarding AI in a more philosophical sense, it has been highlighted that these systems possess the potential to extend beyond mere automated behaviour (Gray & Ertan, 2021).

In my view, the legal terms *mens rea* and *actus reus*, or cognitive and volitional (Donnelly-Lazarov, 2015), when attributed to system functionality, greatly

exaggerate machine autonomy. Autonomy is demonstrated as an inextricable part of system reasoning and actions (Beecher-Monas & Garcia-Rill, 2017). In essence, *mens rea* describes the mental state of committing a crime; in other words, intentionally, knowingly, recklessly, or negligently, and this is vital because it can significantly affect the severity of punishments (Clarkson, 2005).

Actus reus refers to the physical act or conduct that constitutes a criminal offence. The *actus reus* component encompasses the act, results, and circumstances of a specific crime (Herring, 2018). For instance, a war crime can only be committed in a state of war. Furthermore, omission, or the failure to act when there is a legal duty to do so, can also constitute *actus reus*. In simpler terms, to be convicted of most crimes, both *mens rea* (mind or intent) and *actus reus* (the act) must be proven.

In that regard, another way of tackling autonomy is to look at it as bipartite from the perspective of the system's two primary properties. First, I believe that the system's data analytics capabilities can function as cognitive algorithms to facilitate understanding, learning from, and reasoning through the context in which the system operates. That means adapting its behaviour, understanding its function and goal, and anticipating the need for changed behaviour based on the current context. This means dynamically adjusting cognitive engagement levels to match the demand of maintaining operational performance; that is, refraining from situation criticality (Cai & Lin, 2012). A special quality of cognitive machines is expressed when they generate awareness in taking responsive action that reliably matches the characteristics of the case (Koch, 2022).

Second, I agree that it might also be possible that a system serves as volitive assistance that reacts and acts upon the situation, and has the power to execute a course of action. Volition refers to the ability to adapt, implying that the system gains greater control in selecting behaviours. This capacity is likely enhanced when the system operates within a complex and dynamic environment (Chadderdon, 2008). A special quality of volitive machines constitutes the intention to translate the decision-maker into action, which, combined with cognitive properties, enhances the perceptive mind and the active will of a commanding person (Koch, 2022).

In the military domain, cognitive and volitional algorithms in AWS, as I see them, could be conceptualised as deputies to commanders. These algorithms are designed to assist in decision-making processes and, when necessary, autonomously execute actions. This suggests a partnership where military AI complements human command and, under certain conditions, takes the initiative based on directives and learned behaviours, enhancing operational efficiency and responsiveness in combat situations.

Amidst the diverse and often inconsistent nomenclature for autonomous systems (the first caveat), as well as the multitude of hardware-software pairings (the second caveat), the issue of machine autonomy (the third caveat) brings to light three key caveats that begin to outline and influence the discourse of this thesis. In the dissertation, AWS represent the potential of replicating human cognitive and volitive capabilities, making accurate decisions depending on the context, and putting them into action to achieve military goals.

1.1.4 Insight into AI weaponisation

The involvement of autonomous systems in the context of the heightened controversy surrounding AI extends beyond what has already been discussed. To begin with, AI's growing dominance in the defence and security sector might likely lead to an extraordinary milestone in history from which people can gain a lot but lose just as much. For instance, avoidance of human errors and efficiency in decision-making are highly desired; however, the same algorithms can easily cause insecurity in transparency and reliability. The issue is not rogue machines taking over the world but weapons deployed by humans that will drastically reduce our physical security (Russell, 2023).

While AI technologies can perform or take over tasks traditionally executed by human intelligence (van de Poel et al., 2023), it is crucial to acknowledge that our understanding of intelligence itself is not exhaustive, and the precise mechanisms of how the brain functions remain largely elusive. As a result, AI inherits ill-defined intelligence, packed like a suitcase of thinking, cognition, consciousness, and emotion, sporting different meanings in different contexts (Mitchell, 2020).

In one sense, it can be argued that a soldier with a rifle is also an autonomous weapon (**Study III**) because we might never know what went wrong in the use of force for the wrong purposes, resulting in committing atrocities in armed conflict. Algorithmic reasoning is frequently associated with a 'black box' in anticipation of knowing inputs and outputs but not knowing what is happening inside that process. Similarly, it can be argued that one cannot be entirely sure what is going on inside a soldier's mind holding a rifle, a 'black box' of human reasoning if you like, as contemporary neurosciences have no idea of the cortex's internal code (Dubber et al., 2020).

Despite that, no matter how intelligent and autonomous AI agents become in certain respects, at least for the foreseeable future, they probably will remain unconscious machines or special-purpose devices that support humans in specific, complex tasks (Korteling et al., 2021). With this in mind, I have retained military AI as an umbrella term that implies the use of algorithms to replicate intelligent behaviour (Del Monte, 2018) and, at the same time, a technology that elicits a degree of public concern (Horowitz et al., 2023).

I emphasise this point because the terminology related to AI, as discussed in this thesis, tends to be interconnected. Reckoning the caveats I introduced, the interim summary is that autonomous systems pair military AI, and these concepts serve together as a core representation of my studies, as in the following results and discussion chapters. More importantly, military AI qualifies as a blanket meaning for representations among study participants and best characterises how people perceive the adoption of disruptive technologies in the military domain.

1.2 The conceptual basis of the three pillars – ethical, legal and sociological aspects

In connection with the theoretical framework, I think J. J. Bryson nailed it when noting that AI is an adjective, not a noun, allowing us to discover more about how we and societies work and examine the most crucial capacities (Dubber et al., 2020). One of these aptitudes that facilitate making sense of this topic could be the ability to form a mental image of something not completely present yet and to reason about moral issues never before wholly perceived in reality.

From what I have gathered, the implementation of military AI touches a significant part of society, compared to the example of using the internet in the introduction. Studying AI-driven systems in the military domain from distinct positions can comprehensively address warfare's manifold challenges and opportunities and ensure that military AI is understood, developed, and implemented to benefit society. There is a reasonable doubt that autonomous systems might act unpredictably if not designed correctly, primarily operating in real-world environments, which inevitably will affect many fields of life.

Therefore, the multifaceted learning nature of algorithms means no single discipline can fully grasp its implications or potential. An interdisciplinary perspective is increasingly crucial as autonomous systems become more integrated into a military force, especially in terms of unintentionally perpetuating or amplifying biases in the data they are trained on. The findings (**Study I**) became decisive in defining a more concrete framework for theoretical considerations.

The identified areas of concern were closely interlinked, after which grouping into categories resulted in ethical, legal, and, in an extended sense, sociological standpoints (such as the arms race, the threshold of armed conflict, transparency, predictability, trustworthiness, and reliability). As the domain of military AI persistently advances, fresh apprehensions may surface. However, this dissertation primarily focuses on three pillars – normative ethics, legal implications, and sociological challenges associated with military AI applications.

Next, I will introduce deontological and consequentialist views as two main normative theories regarding the ethical pillar that has imparted opposition or congruity to using autonomous weapons (**Study I**). Both perspectives can be complementary when evaluating AI systems from an ethical standpoint. For example, deontology ensures that fundamental principles are not crossed, even if they might produce beneficial outcomes. Conversely, consequentialism can guide decision-making when navigating grey areas (greatest overall benefit) or when principles come into conflict to assess the real-world impacts and make adjustments as needed. Down the line, concerns from the legal pillar are presented from the laws of war perspective, and lastly, the third sociological pillar addresses the conformity of the representation of autonomous weapons in unmanned warfare with ordinary citizens' opinions and experts' attitudes.

1.2.1 Ethical aspects – the first pillar

One of the key concerns surrounding autonomous systems is ensuring that their use is consistent with fundamental moral principles – such as acting to prevent harm to others – that may raise questions about the morality of machine decision-making and the potential for indiscriminate violence without human oversight (Lin et al., 2017). One of the central debates in normative ethics has been between advocates of consequentialism and deontology, in which the former holds that acts are morally right or wrong to the degree that they maximise good consequences (Uhlmann et al., 2009). In light of this, a technological capability only increases over time but ignores that humans are terrible at making rational judgements in dangerous and uncertain situations like combat (Heller, 2023).

Instead of the outcome, deontology holds that certain acts are wrong in themselves, and thus, morally unacceptable means to the pursuit of any ends, even morally admirable or morally obligatory (Singer, 1993). The deontological opposition is based on the belief that autonomous weapons violate fundamental moral principles, such as the duty to respect the inherent dignity of human beings (Lucas et al., 2023). Human dignity is at risk because life-and-death decisions occurring outside human agency or control should be reason enough to impart a solid deontological opposition against using such systems (Saxon, 2021).

It is essential to highlight that moral disagreements draw attention to the inherent subjectivity of moral values that often stem from the diverse ethical perspectives and cultural backgrounds that individuals and groups bring to a discussion. In the realm of military AI, these disagreements become pronounced due to varying beliefs about the ethical implications of using autonomous force in warfare. Furthermore, people differ not only in how they decide to act when confronted with value choices but also in how they justify their choices (Sutrop, 2020).

Ultimately, ethics is contentious regarding judgements of right or wrong on this matter because it exposes unique challenges to the code of conduct surrounding the autonomy of weapons. Moreover, the ethics position has proved to be the overarching theme grounding the respective values and the dominant discourse (**Study I**); therefore, it would be a failure not to notice it. In addition, the potential of algorithms to enhance human operations through AI applications; that is, artificial neural networks, is also discussed in other fields; for example, healthcare (Lenharo, 2023), necessitating ethical validation to ensure that the training of artificial neurons will function reliably within real-world medical, military or other settings.

The dissertation proceeds from the notion that humans can make explicit ethical judgements and generally are competent at justifying them reasonably, regardless of what machines as ethical agents might exist in the future (Anderson & Anderson, 2011). Otherwise stated, the thesis does not aim to explore how to make machines moral agents or how to ensure machines behave ethically but follows the notion of being wiser by making us more aware of the consequences of our actions and be more responsible when deploying such machines (Dubber et al., 2020).

1.2.2 Legal aspects – the second pillar

While moral theories alone are not exhaustive in explaining the algorithmic use of force, a significant weight rests on the legal aspect of how to interpret the applicable law on this matter. As autonomous weapons are designed for the military, the law must be revised accordingly and updated as the technology of war changes (McFarland, 2020). One fundamental debate within this literature deals with the question of whether international humanitarian law (IHL), in its current form, or more broadly, laws of war, sufficiently covers the challenges posed by unmanned military platforms (Christie et al., 2023). The employment of autonomous weapons mainly concerns core principles regulating the use of force: military necessity, humanity, proportionality and distinction (Boothby & Heinegg, 2019).

Notably, most aspects identified in the systematic literature review focused on compliance or non-compliance with these principles, potentially violating or improving adherence to rules that require military force (**Study I**). Following the formulations of IHL, one could argue that AWS will fail to distinguish lawful targets, will cause unnecessary suffering, will lead to excessive collateral damage and fail to set military necessity (Schmitt et al., 2022). I think there are reasonable suspicions that these offences might be realised, especially if machines are trained using biased data, hacked and manipulated for misconduct. However, in parallel, one could argue that AWS lack human weaknesses, making warfare more human because algorithmic targeting is more precise and reduces the risk of harm and if that works that way, we may have a moral obligation to use it (Liao, 2020).

Even though I think there will be no one-size-fits-all standard for using force outside human agency, mainly because deliberate killing can preclude unlawfulness or be justified by understanding the context, such as self-defence by recognising a warning sign of imminent threat. Moreover, a juridical view must consider the context in which problems arise if it adapts to effectively address the dimensions of law and recognises that legal thinking inside society is becoming diversified (Cotterrell, 2018). One might propose that computers could think like humans, using intuition rather than rules, but such speculation is beyond the scope of this thesis. Human cognitive judgement and decision-making are the archetypes of what can be understood as autonomy, but whether machines can adequately replicate such reasoning in situations characteristic of armed conflict environments remains to be solved (Bhuta et al., 2016).

In general, decisions involving lethal force should involve careful consideration and have appropriate safeguards in place to prevent abuses. In particular, the delegation of targeting to machines determined by algorithms poses a specific threat to ascribing responsibility to someone because it might be impossible to identify the moral agents to be held accountable for choices and actions (Liao, 2020). The wake of the obscurity of responsibility could create a legal vacuum in which war crimes happen without actual war criminals.

Therefore, nothing guarantees that AWS can consistently comply with IHL principles, which could create an accountability gap (frequently referred to) in

such cases. We can imagine cases where a human agent does what they can to reasonably foresee potential failure points, but the unpredictable nature of the interaction between the system and environment might lead AWS to engage in an indiscriminate or disproportionate attack (Smith, 2022). Improving legal insight becomes difficult due to the absence of a research object (or even subject) and legal cases involving AWS. This lack of court judgements and legal reviews in terms of “new weapons, means or methods” (AP I, Art 36, 1977) leaves unanswered questions about applying existing laws to autonomy, complicating efforts to grasp legal implications in military use.

Laws of war provide a comprehensive understanding of what core principles demand upholding respect for IHL norms. As I said in the introduction, in the utilisation of AI in the military it is critical to monitor public attitudes, and legal concerns are part of that. The law is an essential and all-pervasive fact of the social condition (Hunt, 1986), and the legitimacy of the law is founded on a broad social consensus – one that can only remain intact if those who give their consent are genuinely informed (Goldstein, 2008).

Qualitatively presented opinions of various concerns and people’s attitudes can have more practical values to make more informed choices in the face of limited knowledge. Because of that, ordinary citizens and military professionals in this study generated valuable insights around the legality of AWS (**Study II, III**). The outcome of discussions could help articulate guiding principles and generate language that highlights the limits and requirements that can be useful in this context (Bo et al., 2022).

1.2.3 Sociological aspects – the third pillar

In this case, the sociological view of current research (**Studies II, III**) also serves as an analytical tool for ethical and legal frameworks to examine how the implications of algorithmic warfare position people’s opinions. The moral and legal thinking of ordinary citizens and experts articulates definition issues surrounding AWS terminology and encourages the highlighting of major issues, such as the delegation of lethal force to machines or algorithmic discretion between right and wrong. On top of that, sociological prospects facilitate communication about moral and regulative outlooks regarding AWS scenarios.

Sociology is concerned with ‘is’ not ‘ought’, with understanding facts, not applying values, but it can clarify much about the role of values in law and society and explain the ways in which these questions arise and the forms they take in certain times and places (Cotterrell, 2018). Several authors point out the hyperbolic nature of the debate surrounding military AI, which seems designed to captivate and alienate the public imagination (Galliot et al., 2021).

That is why social representation theory is the foundation for accurately depicting how study participants think and talk about weaponised AI. According to Serge Moscovici, the theory is based on the contents of everyday thinking and the understanding that a stock of ideas gives coherence to our beliefs, ideas, and the connections that we create as spontaneously as we breathe; essentially, it is a

way of making the unfamiliar familiar to maintain enough flexibility to adjust for differences in groups, cultural matrices, and information circulating in a given society (Moscovici, 2001). That approach also provides a framework for investigating the public's understanding of science and technology, especially if the issue is controversial and can affect many lives (Farr, 1993), which is the case with autonomous systems.

Cumulatively, research on normative ethics becomes critically pertinent when considering autonomous weapons, as it addresses the moral dimensions of delegating decisions to machines. These ethical deliberations influence the laws of war, as introducing AI-driven systems challenges traditional warfare norms. Defining permissible actions and setting boundaries for these systems is a convoluted task, demanding an amalgamation of ethics and legal frameworks. Sociological jurisprudence provides insights into how societies perceive and integrate new technologies into legal frameworks (Cotterrell, 2018), especially those with profound implications like autonomous systems. Laws surrounding such technologies are inextricably linked to society's evolving values and norms concerning technology, warfare, and human responsibility.

In this intricate web of considerations, social representation theory offers a lens for understanding how a given society under the microscope (**Study II**) comes to terms with and conceptualises the role and implications of autonomous systems in warfare. Study participants help to form collective beliefs and shared narratives about these technologies by assimilating new information into existing cognitive frameworks. Thus, understanding these interconnected realms is crucial for shaping societal values, ensuring that laws resonate with the broader ethical and cultural narratives, and navigating the multifaceted dimensions of algorithmic warfare.

Hence, an interdisciplinary understanding encompassing ethics, law and sociological aspects in corresponding themes is vital for the responsible development and deployment of autonomous weapon systems. Within the presented framework, my dissertation aims to supplement this understanding with research on the implications of these aspects and examine the moral dimensions of delegating use-of-force decisions to machines, the challenges of the laws of war in response to these technologies, and societal perceptions and narratives surrounding when AI makes its way into warfare.

2. METHODOLOGICAL CONSIDERATIONS

Considering the complex theoretical background, an exhaustive methodology was essential for delving into the main issues related to military AI. This calls for a foundational understanding of military operations, warfare principles, and familiarity with the autonomous potential used in defence contexts. The implications of military AI are not just technical but also ethical, legal and socio-cultural. For the author, it is clear that a single discipline and method falls short in addressing the nuances of weapons autonomy; therefore, exploring diverse study fields allows for a comprehensive understanding of the challenges and opportunities that autonomy in systems may present.

A structured and systematic literature review approach (**Study I**) utilising academic repositories to find peer-reviewed articles by key authors appeared to be the ideal beginning to study this convoluted theme. The idea of the literature review arose from the need to systematise the current research regarding perceptions of military AI and identify which aspects of autonomous weapons scholars are deemed the most problematic. More notably, there has not previously been a thorough approach to pinpoint the primary concerns related to this intricate subject.

Data gathering for the subsequent studies (**Study II, III**) was conducted in 2020–2022 together with a research group from the Estonian Military Academy (Wagner & Talves, 2023), which explored the social aspects, ethical ramifications, and legal considerations of a project called integrated modular unmanned ground systems (iMUGS). iMUGS was a European Commission project for unmanned platforms to address operational challenges, including enhanced interoperability, increased situational awareness and faster decision-making considering the jurisprudential aspects applicable to robotics, AI and autonomous systems (iMUGS, 2020). To gauge the opinions, perceptions, and beliefs of the general public regarding unmanned vehicles, six focus groups (**Study II**) were established with participants of different education levels and age groups, including both Estonian and Russian speakers. The focus-group study was followed by a series of in-depth interviews with military experts (**Study III**) about the development, presence and importance of autonomous technologies on the battlefield.

2.1 Data and Methods (Study I – Review)

In order to gather the most salient views on autonomous systems in the military, the systematic literature review (**Study I**) generated an excellent overview of the topic, distinguishing the most common concerns and threads to guide and set the foundation for further research. A review protocol was used based on The Web of Science core collection database articles in English from 2015–2020. The refined search criteria applied the title and topic keyword – ‘autonomous weapon

systems'. The eligibility criteria were: (1) the context of warfare and the use of unmanned force in military affairs, and (2) focus on the aspects of autonomous weapon systems and weaponised AI. The exclusion criteria were: (1) full text was not accessible, (2) the systems were merely an illustrative example, and the study did not focus on the autonomy of weapons, and (3) duplicate articles resulting from the title and topic search. In total, 41 publications corresponded after implementing the eligibility and exclusion criteria, a list of which is available in the annexes (**Study I**).

A proportional distribution score showed an average of 6.3 concerns identified per article. The score showed that one-quarter of the 41 studies identified eight or more areas of concern, two of the articles analysed 16 different concerns, and six of the studies discussed a minimum of two concerns. Despite the relatively modest sample size, the evident pattern relating to concerns about systems provides a robust foundation for the research conclusions. The concept of data saturation, wherein additional samples cease to contribute novel insights, bolsters the argument that the existing data set is adequate. Therefore, it is logical to contend that any further data collection would likely reaffirm the initial findings, underscoring the reliability of the preliminary results.

The analytical procedure applied a data extraction form (Petticrew & Roberts, 2006) in order to collect relevant information systematically, pulling out key aspects concerning the discussion of the use of force without direct human intervention. It was imperative that the authors of the articles clearly delineate and link these key aspects; for instance, 'responsibility' for the development, deployment, and utilisation of autonomous systems. Based on their presence, the frequency of the key aspects was gauged on a scale of 0–1, where an aspect had to be discussed at least once in the text to be coded as being present (value '1'). Each additional detected aspect enriched the pool of aspects.

I employed a cross-referencing technique to ensure comprehensive coverage and consistency in my review. This involved revisiting previously reviewed articles to specifically examine whether, for instance, 'responsibility' was addressed in discussions, even if it had not been initially noted. This cross-referencing process was applied systematically across all articles to confirm the presence or absence of this theme, thus allowing for a thorough and accurate representation of the literature review on this aspect.

2.2 Data and Methods (Study II – Focus Groups)

Group discussions are a common methodological means to gather insights about participant perceptions, opinions, and attitudes regarding a specific study (Krueger et al., 2020) and offer an essential means of tracking conversations about developing knowledge and representations when something does become a topic of interest (Davis et al., 1993). Moreover, to gain new ideas and perspectives underlying a particular debate, a focus group can provide a natural environment for participants to both explain and query each other about a particular topic, and

such exchange offers valuable data on the extent of consensus and diversity within a group (Morgan, 1996). Several studies have explored the opinions, perceptions and attitudes toward new technologies through focus groups (Campelo & Katz, 2020; Parent et al., 2000; Vaportzis et al., 2017).

2.2.1 Ethical Matters

The University of Tartu's Research Ethics Committee approved the study, the questionnaire and its annexes, with modifications to minimise mental strain and protect privacy. Participation in the study was rooted in informed consent from individuals; previous binary questions (yes/no) were reframed as open-ended, allowing participants to also voice negative opinions. An outsourced public opinion and marketing research company, Turu-uuringute AS, was responsible for recruiting and instructing the participants, processing personal data, conducting interviews, and anonymising the transcriptions of the study participants.

The data from the participants was stored on a server in a MySQL database, encrypted using SSL, and was only accessible to the panel administrator. It was not shared with third parties or used for other purposes. After the study, the survey company removed the audio recordings and transcriptions from its databases. The survey firm assured the participants that their responses would be kept confidential, and they had the freedom to leave the interview at any point without any repercussions.

The moderators encourage participants to freely share their thoughts as they naturally occur, without the need to search for precise terminology or worry about their lack of expertise in the subject area. Throughout the interview, the moderator reminded them that there are no right or wrong answers, nor are scientifically or technologically accurate responses required, as honest and spontaneous opinions about the topic were valued most. This approach aimed at creating a comfortable and stress-free environment for sharing, and is also described in the research information and consent forms.

2.2.2 Sample, Data Collection and Analysis

Six focus groups were organised, with four groups primarily speaking Estonian and two speaking Russian. The research company translated the Russian-language transcripts into Estonian and provided English translations for all of them. Each group had eight to nine members, culminating in a total of 53 participants, with 25 females and 28 males. Four of the groups had a mix of educational backgrounds, while one was exclusively comprised of those with secondary education and another solely of those with higher education. In all, 33 of the respondents possessed higher education, whereas 20 did not.

The majority of the participants lacked advanced technological knowledge, displaying a general layperson's understanding. The research took place in the spring of 2021 using the Zoom application because of Covid-19 limitations, and

each session spanned roughly two hours. During the interview, members of the research team had the possibility to monitor the process through Zoom's chatroom, but their visual presence was muted (or greyed out) to avoid any disruptions.

The discussed topics were divided into 12 sections of questions and statements designed to elicit responses or counterpoints from participants regarding a specific subject (the questions are provided in **Study II**). As a research team member with extensive experience in the armed forces as a commander, staff officer and legal adviser to HQ, my role involved leveraging my knowledge to shape and deepen our inquiry into autonomous systems. I focused on exploring the military ethics, legal, and operational dimensions of autonomous systems and their potential impact on warfare. My contributions included formulating scenarios that pushed the participants to imagine autonomous technology outcomes. I emphasised developing questions to assess trust in weapons autonomy (**Study II** Q5–Q7), understanding the ethical boundaries, and exploring beliefs through scenarios where autonomous systems operate independently. I aimed to challenge the participants to confront the reality of 'killer robots' and the moral dilemmas they present, such as the accountability for life-and-death decisions made by machines (**Study II**). My input was instrumental in guiding questions in the direction of the potential for a ban on their development (**Study II** Q20) and ensuring that the research encapsulated a comprehensive view of autonomous systems, blending academic insights with operational, real-world considerations from a military standpoint.

The initial two statements sought to measure participant views on robotic technology in general and their perceptions of the surge in technology-driven practices during the pandemic lockdown. After this, participants viewed a brief clip displaying a weaponised unmanned ground vehicle. The video showed the system autonomously engaging a hostile armoured vehicle activated by a visibly remote human operator. Following the footage, participants promptly shared their initial thoughts and emotions regarding the events depicted. Subsequent questions delved into participant opinions on the evolution and implementation of autonomous military technology, facets of artificial decision-making, governance of autonomous systems, and their ethical and legal dimensions.

To explore the participants' viewpoints and thoughts, I conducted a thematic analysis guided by the interpretative reflexive process (Braun & Clarke, 2021). This starts with familiarisation with the data. Initially, I read the transcripts multiple times in English and Estonian to understand the content and context. The participants discuss their emotional reactions, the practical uses of such technology, and ethical considerations. After reading, I marked preliminary codes, for instance, 'human control and moral responsibility'. I systematically coded transcript features relevant to the research question and theoretical framework. For example, 'positive' codes (in favour of the machine's targeting) and 'negative' (opposing the algorithmic targeting process).

The goal was to organise codes and identify broader themes. Thereby, significant insights related to the research question were uncovered. For example, algorithmic military targeting led to the broader theme of decision-making made by

machines and humans, mainly in relation to life-or-death scenarios. At this stage, I named these themes to identify the essence of what each theme could possibly capture in the transcripts. Following such an interpretative reflexive process, I juxtaposed the key aspects of the literature review (**Study I**) with the main themes within the transcripts that encapsulated the complete corpus.

In the final stage, I sought to weave together the analytical narrative with vivid themes to portray the findings of the data in terms of how the multifaceted views of the participants address the research question. By re-examining the themes, I ensured they accurately represent the coded extracts and the entire dataset. Socio-demographic variations between the groups were not considered in the study, as the responses were broadly consistent across the groups. However, the research team noticed that some Russian-speaking participants leaned towards more conservative answers and showed a greater propensity to categorically oppose the development of autonomous weapons.

2.3 Data and Methods (Study III – Interviews)

In the realm of user experience and the tangible implementation of disruptive technology, military officers and enlisted personnel are the primary operators or those directly impacted, making them the end-users. Detailed expert interviews are invaluable for elucidating insights, perspectives, and underlying motivations regarding a particular topic. This approach allows for an in-depth exploration of the respondents' comments, yielding a more nuanced understanding of the issues at hand. These interviews shed light on the rationale behind the benefits or shortcomings of autonomous systems. Delving into particular issues helps gauge the interviewee's unique values, beliefs, experiences and understanding (Scanlan, 2020). Such detailed accounts become a crucial reservoir of knowledge for deeper exploration. The analysis of the interview transcripts is directed at pinpointing the contextual significance of the main topics related to autonomous systems rather than formulating a new theoretical framework on the application of AI-based military force.

2.3.1 Sample, Data Collection and Analysis

Eighteen higher-educated participants participated in the study, including 14 officers, two non-commissioned officers, and two civil service workers. The ranks of the officers spanned from second lieutenant to lieutenant general. The research group actively sought out participants based on specific criteria. Of the 18 interviewees, 10 lacked a comprehensive understanding of robotics, AI, and related systems engineering topics. Meanwhile, eight participants were somewhat familiar with operating a UGV, having either completed operator training or having served in units that utilised UGVs. The opinions of the participants quoted below are labelled with P(x) for practitioners and O(x) for officers. The study operated under the direction of the Defence Forces Chief Directive No. 213 from 28

September 2018, which outlines the coordination and preparation of surveys within the Defence Forces, including the creation of structured questionnaires that address various areas and subjects pertinent to the Defence Forces.

The iMUGS group members conducted the study in Estonian from July to September 2021 as individuals. Seventeen interviews were face-to-face, while one was via a video application. The interview guide comprised 21 primary questions, which were not shared with the interviewees in advance. Except for one, all interviews were recorded. Typically, each session lasted around an hour. Each interview was transcribed, and an anonymous code was assigned to the participant, after which the audio recordings were discarded. All transcriptions have been translated into English, ensuring availability in both languages.

As in the focus-group study, my role involved utilising my expertise to guide the research towards autonomous systems. I focused on inquiries regarding weapons autonomy (**Study III** Q6–Q8), accountability for algorithmic decisions (**Study III** Q10–Q12), and the feasibility of a ban on their development (**Study III** Q17–Q20) to ensure that my findings from the academic literature review merged with the perspective of military professionals.

The questions were open-ended, prompting the interviewees to express themselves without reservation. The interviewers, when deemed necessary, requested clarifications or further details from the respondents and also posed supplementary questions to stimulate discussion. This method enabled a deeper exploration of the topics addressed. The terms ‘AI’ and ‘autonomy’ were intentionally left undefined for the participants to avoid any preconceived notions or biases that might obstruct the dialogue. Participants were encouraged to treat them as synonymous, particularly in the context of describing unmanned warfare within the military sector.

The goals of content analysis are to examine statements to draw conclusions by pinpointing distinct textual features (Titscher et al., 2000). In this scenario, the statements reflect end-user viewpoints, emphasising the challenges with using AI in the military sector. More specifically, the qualitative method identifies the presence or absence of certain content attributes as indicators for categorisation (Krippendorff & Bock, 2008). Moreover, to deepen the participants’ views and interpretations, the transcribed texts were analysed using a two-step qualitative content analysis method: deductive category application followed by inductive category formation (Mayring, 2000).

I initially established categories deductively based on the comprehensive literature review encompassing the perspectives of various scholars on autonomous weapons (**Study I**). This approach ensured that the most prominent elements of the review were deductively integrated to reinforce and broaden the current understanding of military AI. The inductive category formulation was receptive to novel manifestations and actively sought patterns that could modify or even challenge the deductive categories related to autonomous systems. Delving into the contextual significance of the participants’ attitudes, perspectives, and convictions concerning autonomous weapons shaped the central themes of the research.

2.4 Data Visualisation in the Findings

To synthesise and present the findings from the systematic literature review, I employed R-Studio, an environment for statistical computing and graphics, leveraging the ggplot2 package (Kronthaler & Zöllner, 2021). The dataset was compiled from the review, with each entry representing a distinct aspect identified within the corpus. These aspects were quantified based on their frequency of occurrence across various articles. The visualisation was operationalised through a bar plot chosen for its straightforward interpretation and ability to communicate data distribution effectively. The plot was designed to display each aspect along the x-axis, with the height of the bars on the y-axis indicating the number of occurrences. This design allowed for immediate visual identification of the most prevalent themes within the literature. In terms of enhancing the plot's interpretability, horizontal lines were introduced to represent the median and mean values of occurrences. These reference lines provide a benchmark against which the significance of each aspect's frequency could be evaluated.

2.4.1 Similarity Analysis of Qualitative Data

To uncover underlying patterns and themes within the qualitative dataset, which encompasses focus-group discussions and in-depth interviews with military experts, I utilised the computational capabilities of Iramuteq (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*). This analytical approach was first introduced to our team by Prof. Wolfgang Wagner, whose guidance was instrumental in its application. The effectiveness of this method is evident in its successful use in a co-authored article by our research group, where it provided insightful data visualisation and interpretation (Wagner et al., 2023).

This statistical software tool interfaces with R to perform a similarity analysis on the collected transcripts, enabling a detailed exploration of the textual data (Chaves et al., 2017). This methodological approach made an intricate assessment possible of the interrelationships between the expressed ideas, themes, and concepts within the transcripts of the focus groups and expert interviews. The process entailed removing extraneous elements and segmenting text into analysable units, setting the stage for accurate similarity analysis.

The similarity analysis conducted via Iramuteq facilitated a comparative analysis, revealing the associative relationships between thematic elements from the multifaceted conversations with military experts and focus group participants. By generating a similarity matrix, Iramuteq evaluated the co-occurrence of keywords and phrases, thereby quantifying the strength of association between them. The graphical outputs visually synthesised the thematic content from both datasets. These similarity graphs represent significant themes (or *nodes* in the software) extracted from the data, with the edges reflecting the frequency and strength of the connections between them. The edges (or connections) between nodes would indicate the degree of similarity or relatedness between them.

The deployment of similarity analysis furnished a perspective on the thematic constitution of a detailed thematic map that illustrated commonalities and diver-

gences in discussions around military subjects. The insights gleaned from this analysis enriched my understanding of individual and collective viewpoints and underscored the nuanced complexity inherent in military AI discourse.

To support the visual analysis based on Iramuteq's corpus, I employed R-studio (Posit team, 2023) statistical tests for the emerging themes, which provided a mathematical backbone to the visual representation. I decided to analyse themes of co-occurrences in a particular cluster and utilise Spearman's rank correlation coefficient rather than Pearson's correlation coefficient due to the non-normal distribution of our data, as evidenced by the Shapiro-Wilk test results and the presence of non-linear relationships between variables, making Spearman's coefficient the more robust and appropriate measure for assessing the monotonic relationships within the dataset.

For instance, the themes 'human' and 'robot' in Figure 2 (see 3.2.1) according to Iramuteq provide a Spearman's coefficient in R-studio of 0.235, with a p-value of about 0.073. This score indicates that while there is a weak positive correlation, that relationship is not statistically significant based on the standard level of 0.05. The combined co-occurrence in the dataset is 38, which indicates a notable level of joint discussion across the corpus.

In sum, the linkage is not strong enough to suggest that participants frequently tended to associate between themes; instead, there is a degree of caution in drawing parallels to robots. However, a thorough qualitative analysis of the discussions is necessary to support a more definitive conclusion about the participants' attitudes and perceptions. For that reason, I decided to integrate qualitative and quantitative findings to reinforce the analysis of vital aspects of AWS to thoroughly examine the data from multiple perspectives.

It has also been argued that particular expressions and their statistically significant patterns across texts inform notions of an individual repertoire (Bednarek & Martin, 2010). Interpretative repertoires provide flexible resources that enable participants to discuss topics creatively and are instrumental in understanding how the research object is constructed (Keller & Kalmus, 2009). Repertoires in discursive psychology are defined as broadly discernible clusters of terms, descriptions, and figures of speech assembled around vivid images (Wetherell & Potter, 1992). In that vein, I propose that representations of nodes serve as diverse repertoires of participants, in which clusters reveal the range of discursive and thematic emphasis to elucidate connections within affiliations and distinct features, in which participants navigate and negotiate the topic, thereby enriching our understanding of the discourse.

Employing these specific methods made a nuanced analysis of the transcripts possible, revealing intricate patterns and relationships that might have been overlooked with other analytical approaches. However, it is essential to acknowledge the limitations inherent in any methodological approach, including the potential for overlooking subtler, less statistically prominent themes. Despite these limitations, this methodology provided an understanding of the attitudes and experiences of the focus groups and military personnel regarding the subject matter.

3. FINDINGS

This dissertation's first objective was to identify the main concerns associated with introducing autonomy to the battlefield (**RQ1**). The literature review in **Study I** stems from the necessity to organise existing research on perceptions of AWS and pinpoint the elements scholars find most troubling. The necessity for the first study arose from the lack of comprehensive data available to effectively identify and address key issues in the development of autonomous systems. Consequently, initiating a more holistic discourse on the implications of unmanned warfare hinges on discerning the most frequent issues and uncovering a unifying theme that can steer subsequent inquiries.

Study II aims to meticulously address the main aspects resulting from the literature review, such as the characteristics of decision-making by machines compared to humans, particularly in life-or-death scenarios, and seeks to understand the intricate dynamics involved. The study also explores the issue of the accountability gap for war crimes potentially committed by autonomous weapons. Furthermore, it critically examines the debates concerning the development of autonomous systems, weighing the moral justifications for either endorsing their use or exerting pressure for an outright ban. The central aim (**RQ2**) was to investigate the defining attributes of autonomous systems used in military applications and elucidate the concept of weapon autonomy.

The last study aims to uncover additional reasoning that supports the issues highlighted in previous studies. In examining the complexities of weaponised artificial intelligence, **Study III** seeks to uncover the dilemmas that emerge within this new era, to understand the factors contributing to uncertainties surrounding algorithmic decision-making in autonomous systems, and to consider strategies the military might employ to manage the unpredictable elements of unmanned warfare (**RQ3**).

This chapter employs a series of illustrative figures to meticulously depict the principal findings extracted from the literature review, concentrating on the key aspects of military autonomy. These visual representations are intended to facilitate a concise comparison between the academic literature and the empirical data gathered from sequential focus groups (**Study II**) and in-depth interviews (**Study III**). This approach has been adopted to augment the clarity and comprehensibility of the findings presented in the dissertation studies. It serves to highlight both the congruence and divergence between the established theoretical framework and the practical insights gleaned from the qualitative research methods.

3.1 A Compendium of Arguments (Study I)

Figure 1, created in R-studio, illustrates that discussions predominantly revolve around the fundamental principles of international humanitarian law (IHL), which range from individual principles to the analysis of the entire set of principles. The main principles of IHL are the distinction of lawful targets, the proportionality demanding civilian harm not to be excessive for military advantage, and military necessity permitting an engagement when an action results in civilian harm (Fleck, 2021). Concerns regarding AWS largely stem from debates about their ability or inability to conform to IHL principles. Critics express significant scepticism, arguing that the complex judgements required to uphold these principles are currently unattainable for weapons, given that machines cannot yet surpass their creators in such nuanced decision-making. On the flip side, a faction of contrarians posits that AI-driven systems could potentially execute IHL principles more reliably, unimpeded by emotional responses such as panic or distress that might compromise human decision-making in warfare.

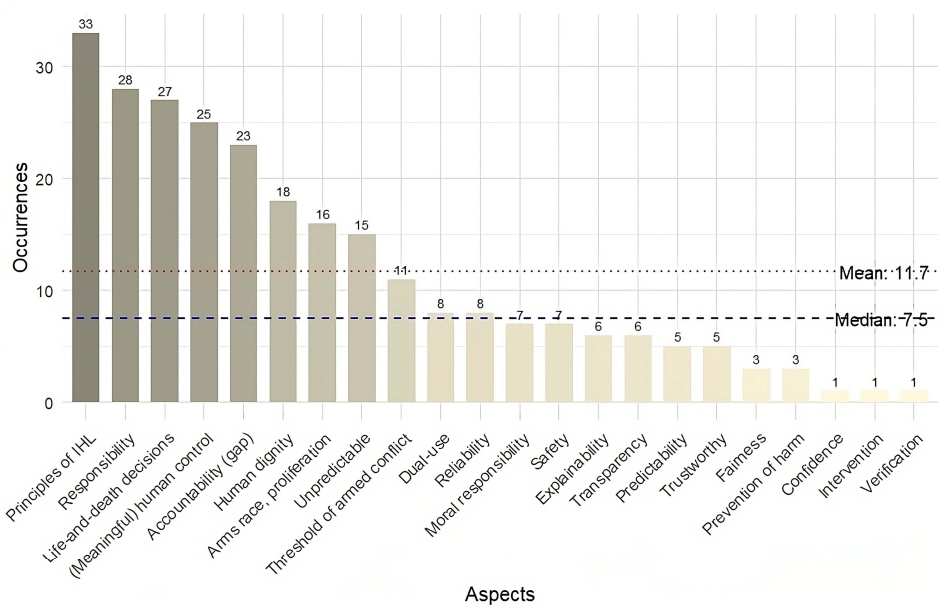


Figure 1. Twenty-two aspects identified in the literature review from a total of 41 articles. R-studio.

Figure 1 suggests thematic clusters where certain aspects, such as ‘responsibility’, ‘accountability gap’, and ‘moral responsibility’, frequently overlap, pointing to a network of interrelated concerns. This interconnectivity informed me of possible categorisation, highlighting the dominance of themes and aiding in synthesising the insights from the literature. Based on Figure 1 and according to my interpretation, in addition to IHL principles, there are prominent recurrent themes such as matters of responsibility, dehumanisation in targeting, the imperative of human

oversight, and the proliferation of unmanned warfare. Less frequently, numerous aspects carry significant ethical or technical weight in their own right, contributing to discussions surrounding autonomous weaponry and highlighting its importance in academic debates.

As a further matter, the theoretical scrutiny of autonomy in weapons can be organised under various normative ethics traditions. However, a number of these discussions about worrisome aspects lack a foundational theoretical framework. About half of the considerations did not anchor their arguments in a particular theory of ethics, while the rest were framed within ethical theoretical boundaries. The literature review indicated that the other half of the considerations mainly focused on deontological principles over those of utilitarianism and virtue ethics. Although the relationship between liability, compliance with IHL, and ethics was not always explicitly made, the considerations often implicitly resonated with deontological ethics. Many scholars have contended that the very act of machines making life-and-death decisions in autonomous mode is a compelling deontological argument against the deployment of AI-enabled platforms.

3.2 A focus-group perspective (Study II)

A similarity analysis was derived from the focus group transcripts using Iramuteq and a binomial score, which quantifies the strength of association between two nodes based on the presence or absence of corresponding attributes. I employed the Kamada-Kawai layout algorithm to visualise the complex interrelations within the data. This approach is particularly adept at spatially positioning nodes to reflect the strength of their mutual associations, resulting in a graph that intuitively maps the relational dynamics between key thematic elements (Kamada & Kawai, 1989).

Similarity analysis is a well-suited method for uncovering patterns and relationships in qualitative responses. Analysis parameters determine the significance of an association for identifying connections in responses, ensuring statistically significant patterns (Ratinaud, 2023). Figure 2 conjures a picture of relational networks and visually maps how specific ideas are interconnected. This visualisation supports qualitative content analysis in revealing how specific themes or topics are grouped in the discussions, indicating shared or divergent perspectives among the participants.

In Figure 2, each cluster represents key themes derived from the interview transcripts, and the edges signify the associative relationships between these themes based on the frequency of their co-occurrence. The edges between nodes in these clusters indicate the strength and frequency of associations, with thicker or more numerous edges suggesting stronger or more commonly expressed linkages, as the graph provides a macro-level view of the corpus context.

3.2.1 Distinguishing Human and Machine Decision-Making

Although the study did not specifically ask for a definition of an autonomous weapon system, it became clear from various responses that participants generally perceive it as an autonomous system designed for military operations, with the ability to decide when to apply force to a certain extent independently. However, there was consensus among the focus group that despite the potential of machines to augment human abilities, they should not be relied upon to make autonomous decisions.

A notable finding in Figure 2 was the strong edge between the nodes labelled ‘person’ and ‘machine’, pointing to the fact that these linked nodes are frequently discussed in conjunction and provide support for the salience of this topic among focus-group dialogues. The prominence of the ‘person’ and ‘machine’ nodes is indicated by the larger size of the text, reflecting that these themes occur more frequently or hold more significance within the context of autonomous systems paradigms.

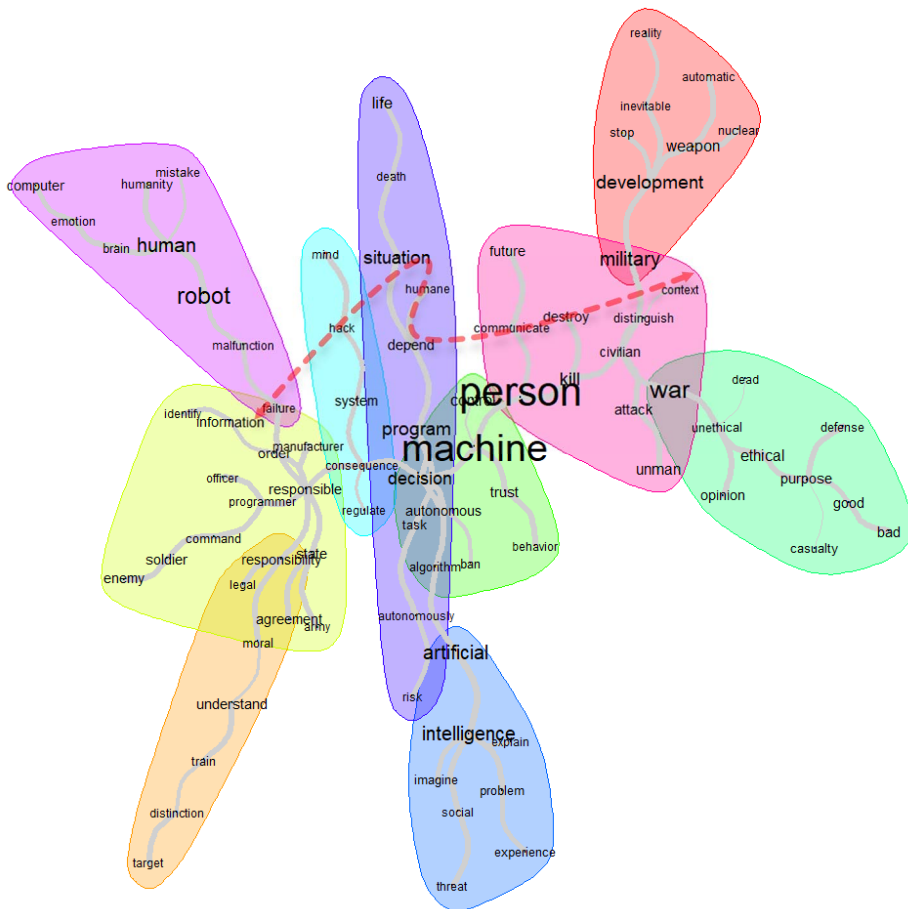


Figure 2. Similarity Analysis of Focus Group Transcripts. Iramuteq.

The intricate nuances of opinions surrounding the context of decision-making processes can be seen on the person-machine node in the centre. The edges connecting these nodes indicate the comparative quality of the ‘decision’. This caution was rooted in the inherent complexity of high-stakes decisions reflected by the complex conundrum of grey areas that must be considered. These nuances made participants suggest whether a decision made by a machine or a person is better or worse depending on the specific situation and the information available, requiring an interpretation deeply dependent on the context. In Figure 2, a red dotted curved line indicates and guides attention to decision-making aspects for the visual representation.

Table 1 displays the statistical analysis for the theme ‘decision’ in relation to the cluster’s themes. Notably, ‘person’, ‘depend’, and ‘situation’ exhibit moderate but statistically significant correlations with ‘decision’. In addition, a pronounced co-occurrence with ‘machine’ and ‘person’ underscores the ongoing discourse regarding the nature of decision-making across distinct forms of intelligence.

Table 1. ‘Decision’ node statistics. R-studio

Theme	algorithm	autonomous	consequence	depend	kill	machine	person	situation	task	trust
Correlation	0.22	0.26	0.30	0.37	0.11	0.29	0.37	0.40	0.23	0.09
p-value	0.091	0.046	0.020	0.004	0.393	0.027	0.004	0.002	0.074	0.495
Co-occurrence	11	15	7	23	35	92	90	44	14	17

Participants expressed concern that, while machines can process information and execute tasks with impressive speed and efficiency, the subtleties of ethical judgement and emotional intelligence, crucial to decision-making in chaotic and unpredictable environments, are domains still dominated by humans. The discussions highlighted a general belief that human judgement, despite its fallibility, incorporates a depth of understanding and a capacity for moral reasoning that machines have yet to achieve or replicate.

I noticed that while there is an appreciation of the potential of machine-based decision-making in structured and well-defined scenarios, there is considerable apprehension about its use in situations where decisions are fraught with ethical dilemmas and require a nuanced assessment of complex variables. The insights from the participants highlight the complex difficulties involved in creating autonomous systems that can match human intuition and ethical judgement, particularly in the fast-paced and unpredictable environment of military operations where mistakes can have severe consequences. They stress the importance of retaining human oversight for critical decisions, especially those concerning life and death, suggesting that such grave responsibilities should not be transferred to machines.

3.2.2 Assigning Responsibility Involving Autonomous Weapons

Considering legal liability, the yellow cluster on the bottom left of Figure 2 represents the edges of ‘responsible’ when using autonomous platforms. The owner of the weapons system was deemed accountable, indicating that liability rests with the state’s military forces rather than the system’s manufacturer. In Table 2, ‘responsible’ has positive correlations with all themes in the cluster, and notably strong correlations with ‘programmer’ (0.69), ‘order’ (0.63) and ‘officer’ (0.60). The coefficients for ‘command’ (0.58), ‘state’ (0.35), and ‘soldier’ (0.27) exhibit lower correlations but are still significant based on p-values.

Table 2. ‘Responsible’ node statistics. R-studio.

Theme	Correlation	p-value	Co-occurrence
army	0.32	0.014	11
command	0.58	0.000	4
manufacturer	0.59	0.000	8
officer	0.60	0.000	17
order	0.63	0.000	16
programmer	0.69	0.000	16
soldier	0.27	0.035	5
state	0.35	0.007	5

The lowest p-values <0.0001 for the pairs with ‘order’ and ‘programmer’ indicate that the correlations are unlikely to be due to random chance. Although statistics indicate that ‘order’ and ‘programmer’ frequently co-occur in the context of responsibility, this does not equate to their identical significance in attributing accountability. Content analysis revealed a nuanced perspective on responsibility: while individuals issuing orders were widely regarded as accountable, programmers, on the other hand, were not similarly held liable for the use of autonomous weapons.

Within this accountability framework, I have complemented the statistical findings with outcomes from the qualitative analysis of the focus group transcripts (**Study II**). Military commanders, including the head of defence, were identified or pinpointed as the individuals who issued direct commands to their subordinates. As the group discussions progressed, a shared viewpoint emerged that a soldier or a machine, or anyone outside the command structure, merely serves to carry out these orders. The prevailing sentiment among the participants was that the person within the military ranks who instructs the use of an autonomous weapon bears the responsibility for the outcomes of such an order. The participants of the focus-group study expressed a shared concern about the potential for a lack of clear accountability regarding war crimes committed by autonomous weapons. They were firmly convinced that a responsibility gap is unacceptable and insisted that it is essential for individuals to be held accountable for any crimes perpetrated by an AWS.

3.2.3 Arguments for and Against the Development of Autonomous Weapons

In Figure 2, the red cluster in the top right illustrates edges concerning the development of military systems autonomy. When queried about their views on the widespread advancement of military AI and whether they believed the concept of AWS presented an ethical dilemma, the responses from the participants were unexpectedly nuanced. Rather than providing straightforward affirmations or objections, discussions often evolved into intricate arguments concerning the general ethics of autonomous systems. Participants noted that ethical judgements rely heavily on context, accurate information, and primarily on the intended use, asserting that machines cannot be inherently labelled as ethical or unethical.

Table 3. ‘Development’ node statistics. R-studio.

Node	Correlation	p-value	Co-occurrence
bad	0.04	0.781	22
ethical	0.24	0.068	22
good	0.31	0.018	31
inevitable	0.36	0.005	6
purpose	0.27	0.040	25
reality	0.35	0.007	6
stop	0.15	0.261	10
unethical	0.41	0.001	13

There was no explicit agreement among the groups regarding what constitutes ethical behaviour or how machines might emulate human ethical reasoning. As indicated in Table 3, the ‘ethical’ and ‘unethical’ statistics suggest the conditional nature inherent in the ‘purpose’ when giving the meaning of ‘good’ vs ‘bad’ regarding development. Concerns were raised that the adoption of AI and autonomous machines might lead to a loss of human-centric thinking, prompting people to behave more mechanically. The study clarified that ethical considerations are paramount to the participants and that the creation of autonomous weaponry must be morally defensible. Nonetheless, the feedback from the focus groups suggested that the development of military autonomous systems is not considered inherently immoral, but its ethical standing is dependent mainly on the intended application.

With respect to the development process, there is ambiguity among the participants about whether to prohibit autonomous systems completely or to regulate their advancement cautiously. It was suggested that technology could be acceptable in scenarios where it does not directly endanger human lives. The consensus was that the outright ban of AWS was unlikely, and their development was ‘inevitable’ as supported by the statistics ($r=0.36$; $p=0.005$; $n=31$) and not entirely favoured, as their development could have positive applications in civilian and military domains.

The trajectory of autonomous weapons systems seems to echo the historical path of nuclear weapons as it stands out in the development cluster. Despite international efforts to regulate nuclear arms, their existence and development have been a persistent reality (SIPRI, 2021), which suggests that a ban on AWS might face similar challenges. Technological advancement is often unstoppable, particularly in areas with significant military applications. The question then becomes not if but how such technologies should be governed.

The central reservation about weaponised AI was their significant risk to human life and the unpredictability of their advancement. Despite scepticism about whether regulations could mitigate these risks, there was a general resignation to the idea that the progress of AWS in military use is bound to happen. As a result, there was collective agreement on implementing strict oversight and standards for weaponised AI research. The participants concluded that any progress in this field should be cautiously approached and that international consensus on AWS regulation is essential.

3.2.4 Discomfiture over Weaponised AI

The study revealed a complex attitude towards autonomous weapons in the Estonian armed forces. Participants generally favoured using autonomous systems over troops for national defence, citing the potential to save lives and reduce the emotional toll of losing soldiers. However, some expressed reservations about the operational capabilities and moral aspects of autonomy in weapons, noting the irreplaceable value of human will and patriotism in defence. Interestingly, participants preferred to face human soldiers rather than enemy robots when considering the opposing side, referring to the unpredictability and the intimidating aspect of advanced technology. This preference was echoed, for instance, by respondents at the World Economic Forum (Del Monte, 2018). Overall, the study highlighted a preference for using autonomous systems for one's defence while facing human soldiers in opposition, presenting an ethical dilemma and the paradoxical nature of these preferences in the context of advancing AI technology in the military domain.

3.3 Insights from Military Experts (Study III)

The analysis found mixed attitudes among military personnel towards the autonomy of weapons. Despite acknowledging the inevitable progression of autonomous systems in the military, scepticism about their implementation was prevalent, especially among senior officers. Participants questioned the technology's reliability and practicality and the need for a significant shift in military mindset and training to integrate such emerging technology effectively. Thinking of autonomous systems led to varied opinions: those with hands-on experience were more optimistic about their potential, while others doubted their readiness for military use.

The ambiguity of integrating autonomy exists in the multi-domain operational environment, which might impact effective and coherent command and control (C2), enabling commanders to effectively convey their intentions and exert their authority (Granillo, 2021). Concerns about multi-domain operations were common, with many feeling that autonomy would be premature as a substitute or replica for military functions in joint operations. The consensus was that autonomous systems might demand more extensive human resources than traditional armed forces.

From one perspective, the swift evolution of AI applications is happening at a speed surpassing the capacity of armed forces to cope effectively. Conversely, considering the Estonian army's heavy reliance on reservists, it became evident that there would be a significant challenge in keeping up with technological advancements in military training. The following generalisation can be drawn: in integrating military AI within armed forces reliant on reservists, small countries similar to Estonia should focus on creating modular, role-specific AI training programmes that do not require in-depth technical expertise. Moreover, it is advisable to incorporate civilian AI expertise through reservists, enabling easier adoption during defence readiness exercises.

Overall, there was a sense that AI and autonomous technology are often over-emphasised, overshadowing the fundamental role of soldiers in conventional warfare. The preference was for investing in conventional military equipment over expensive high-tech systems because, until now, soldiers engage in combat, not weaponised AI systems.

3.3.1 Military applications of warfare in autonomous mode

I employed the Iramuteq software in a manner consistent with the previous approach to visualise the transcripts of the interviews with military personnel. In Figure 3, the green cluster around 'person' symbolises the centrality of the human role and judgement in this discussion. The 'person' cluster in relation to nodes on the left indicates a close association between the themes of military leadership and responsibility, which were prominently discussed. That also might reflect conversations about the role of commander decision-making, leadership qualities, and personal accountability for actions in military operations.

The cluster of 'war' on the right in dark pink, located between 'person' and 'machine', emphasises the inherent risks and mistakes posed by robots making command decisions. The individual's role underscores the indispensable value of human empathy and moral reasoning in contrast to the cold, algorithmic logic of robots. Such a comparison not only illuminates the stark differences between human and machine decision-making but also calls for urgent and comprehensive regulations to ensure that the emerging autonomy of machines does not compromise the sanctity of human life. This argument solidifies the stance that in matters of life and death, the human mental and ethical framework should remain paramount over autonomous systems, regardless of their perceived efficiency or rationality.

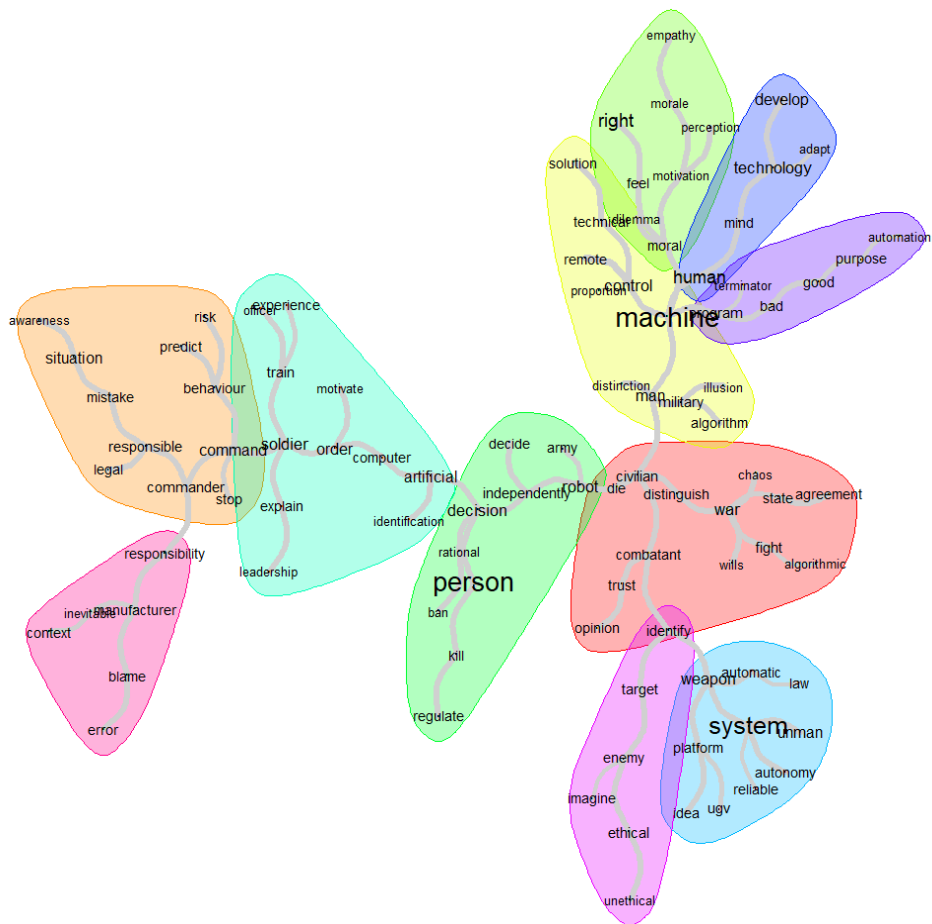


Figure 3. Similarity Analysis of Interview Transcript. Iramuteq.

The cluster around ‘machine’ is visually much more nuanced than the ‘person’ cluster, highlighting the ethical domain of integrating machines in warfare. The nodes associated with ‘machine’, ‘human’, ‘good’, ‘bad’, and ‘develop’ suggest the debate on the moral dilemmas posed by AI in military contexts. Undoubtedly, the node ‘terminator’ in the purple cluster carries significant symbolic weight; however, this association stems from experts’ thoughts inspired by the iconic representation in the movie rather than from a realistic concern. The compliance of the IHL distinction, proportion, and necessity (purpose) principle raises crucial moral and practical concerns. A central issue is the machine’s ability to identify military targets, a task fraught with complexity and uncertainty, especially in urban warfare settings (**Study III**).

The identification challenge is complicated by tactics such as fighters blending in with civilians or using camouflage face colours, rendering even sophisticated facial recognition technologies ineffective. The near-unanimous opinion in the study underscores a deep scepticism about the reliability of AI in

making critical distinctions on the battlefield. This sentiment reflects a severe concern about the inherent limitations of AI in understanding the nuanced and often ambiguous nature of modern warfare, where the lines between combatants and non-combatants are frequently blurred.

In contrast, it is also argued that human soldiers, burdened by fatigue, stress, and other human limitations, are also prone to errors in judgement under extreme conditions. The study suggests that an autonomous system, devoid of human frailties, might be more reliable in identifying targets by focusing on anomalies a human might overlook. This perspective does not entirely promote the utility of AI in military operations but instead advocates for a cautious approach, recognising both the potential and the limitations of autonomous systems. A recurring theme is the moral imperative and that such decisions must ultimately remain within human control. This view emphasises the ethical necessity of keeping the final decision-making power in human hands, reflecting a deep-seated conviction about the moral responsibility and agency inherent in warfare.

The cluster around ‘soldier’ and ‘command’ focuses on the traditional aspects of military operations, including command structure, training, decision-making and responsibility. The linkage between these nodes indicates discussions centred on the command chain and the implications of introducing AI within this structure. The question of machine autonomy in military decision-making centres on the quality and effectiveness of such decisions. Generally, there was consensus among the experts that machine decision-making is problematic, if not impossible, especially under the dynamic conditions of warfare, where struggling with such conditions is unavoidable even among experienced commanders.

For instance, while securing a position might be more effectively managed by a robot, the nuanced decision to use minimal force cannot be distilled into simple rules for an autonomous system. That highlights the context-specific nature of military decisions that machines cannot replicate at the moment. Moreover, the linkage to the program node highlights the judgement in differentiating between good and bad and brings a case against automation. The sentiment in the study is that AI, despite its advancements, remains a program bound by its coded rules and incapable of replicating nuanced human understanding and adaptability.

The potential impact of autonomous systems on the art of warfare is acknowledged by military personnel. For instance, commanding machines could be more straightforward, as unmanned platforms do not require motivation or deal with fear and emotions like humans. However, employing autonomous technology would not fundamentally alter the essence of warfare, which essentially is viewed as a struggle of human will. The point was made that machines, lacking freedom of will, cannot change this intrinsic aspect of warfare, suggesting that the role of autonomous systems might be more supportive than transformative. Despite the potential for machines to execute commands and assist in operations, the ultimate decision-making power and trust were still firmly placed in the commander’s hands. Trust in human decision-makers over machines highlighted a deep-rooted belief in the superiority of human judgement and intuition in warfare.

Nodes around ‘command’ in Table 4 are evidential regarding C2 because soldiers need leadership, a behaviour that is very difficult to expect from AI. For instance, the statistics for ‘commander’ indicate a positive, strong and significant correlation ($r=0.67$, $p=0.002$) with the ‘command’ cluster with relatively high co-occurrences ($n=52$), while ‘artificial’ shows a more modest and insignificant correlation ($r=.44$; $p=0.065$) and somewhat lower co-occurrence ($n=48$).

Table 4. ‘Command’ node statistics. R-studio.

	artificial	awareness	behaviour	commander	experience	explain	identification	officer	order	responsible	risk	soldier	train
Correlation	0.44	0.15	0.68	0.67	0.35	0.41	0.01	0.35	0.44	0.21	0.18	0.46	0.13
p-value	0.065	0.557	0.002	0.002	0.157	0.092	0.962	0.155	0.067	0.408	0.486	0.056	0.618
Co-occurrence	48	7	16	52	35	11	6	7	52	44	22	67	36

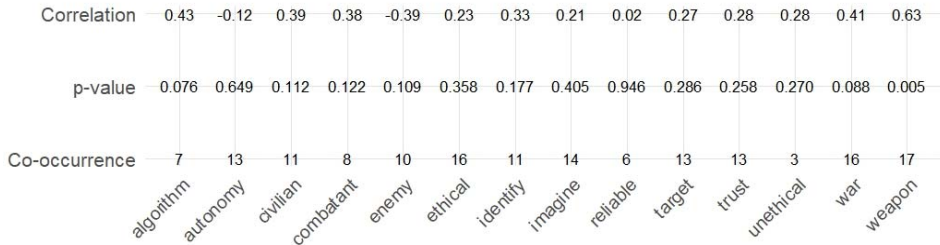
The correlation ($r=0.21$) and co-occurrence ($n=44$) of the node ‘responsible’ with ‘command’ denotes conversations about the accountability and expectations placed on military officers. In other words, a decision-maker who must navigate complex situations, often under significant pressure, has substantial implications for subordinates who require leadership, uphold ethical standards, and exert authority within military operations. Taking into account what has been said in terms of C2, it is imperative to acknowledge that the linkage between clusters on the left points to a rich dialogue about taking responsibility in the age of increasing autonomy, which is seen to be borne by the commander giving an order and no one else.

3.3.2 The Unknown in Unmanned Warfare

The intersection of the ‘war’ and ‘system’ clusters presents a compelling narrative on the role of autonomous systems in the contemporary battlespace. The positive correlations of themes such as ‘algorithm’ ($r=0.43$), ‘civilian’ ($r=0.39$), ‘combatant’ ($r=0.38$), ‘identify’ ($r=0.33$) and ‘trust’ ($r=0.28$) encapsulate the essential and stark realities of armed conflict, concentrating on the autonomous targeting whose reliability becomes a grim necessity in the dynamics of combat operations. Adjacent to the ‘system’ cluster, depicted in violet, are the themes ‘target’, ‘ethical’, and ‘unethical’, suggesting proximity to the moral attributes in the targeting cycle of autonomous systems. The associated linkages in Table 5 regarding ‘ethical’ ($r=0.23$; $n=16$) or ‘unethical’ ($r=0.28$; $n=3$) suggest how emerging platforms may confront the future of military activities and warfare. Figure 3 highlights the imaginative consideration of the potential of systems, balanced against the deliberation of ethical/unethical in identifying targets autonomously. The edges connecting these two, plus the ‘war’ cluster, are significant as they represent the

roles on the battlefield, including delegating targeting to autonomous systems, oversight, and the trust placed in algorithms.

Table 5. ‘System’ node statistics. R-studio.



Introducing autonomous systems might necessitate more technical staff and legal advisers to interpret situations and advise on the technicalities and legality of actions, potentially complicating the decision-making process rather than simplifying it (**Study III**). In particular, concern was raised about maintaining situational awareness and ensuring coherent operations between manned and unmanned units in the same battlespace. In a similar vein, it is argued that moral deskilling is an additional risk with automation to act morally (Moelker & Schenk, 2018). The moral component of warfare, the soldiers’ inner will to fight, and the patriotic mentality were seen as crucial elements missing in machines. It is argued that there is a risk of moral deskilling. Machines, operating purely on algorithmic calculations, cannot replicate the fighting spirit and morale of human units, potentially underestimating the unit’s combat performance. While machines can be programmed for specific scenarios, the unique human ability to share and learn from experiences remains unmatched.

In the realm of the unknown in unmanned strategy, there is a growing belief among experts that the future of warfare will increasingly involve algorithmic rivalry, where AI-embedded systems play a crucial role in deceiving and outmanoeuvring the opponent’s AI. They anticipate that future systems will be designed not to counter adversary human resources but to predict and counteract the patterns and connections formulated by rival military AI. This approach stems from the belief that a computer can more effectively anticipate the strategies of other AI entities than predict human behaviour.

Training an autonomous platform in military ethics and ROE (rules of engagement) is seen as a daunting, if not impossible, task. That concern was primarily rooted in the belief that autonomous systems cannot achieve a level of autonomy that would warrant attributing ethical considerations to their actions. The study indicates a prevailing view that ethical categories are not applicable or relevant to the actions of machines, stressing that ethical frameworks apply to human actions in warfare. The consensus was that war, by its nature, tends to negate morality, arguing that the ultimate goal in warfare is to defeat the enemy.

The interviews did not highlight military ethics as a central concern in using autonomous systems. This outcome might be due to the approach used in the interviews, which possibly failed to prompt more profound reflections on ethical matters. It seems that the practical aspects of using military AI in warfare overshadowed the ethical considerations, reflecting a more pragmatic approach to warfare. Nevertheless, the discussions provided valuable insights into the unknown behind adopting autonomy in weapon systems and the contextual nuances shaping perspectives on this issue.

3.4 A Convergence of Thoughts: Citizen and Expert Insights on Autonomous Weaponry

The similarity analysis conducted by Iramuteq signalled that the clusters ‘person’ and ‘machine’ emerge as focal points (indicated by the larger text), revealing the central confrontation in the discourse. The peripheries of these clusters are rich with terms that articulate the integrated edges between human and artificial agents. The themes ‘decision’, ‘responsible’, and ‘develop’ stand out, suggesting a pronounced focus within the studied discourses.

Both in Figures 2 and 3, ‘decision’ appears as a significant theme, indicated by its central positioning, suggesting that decision-making is a crucial topic in the discourses in both studies. This is particularly expressive by reflecting on how artificial entities might interact with and affect the essence of being human, especially where ‘kill’ and ‘death’ discretions are paramount. The discourses in the two studies are also linked through nodes like ‘autonomously’ and ‘independently’, implying that attention revolved around trust in artificial judgement.

Focus-group participants anticipated challenges in replicating human-like ethical reasoning and intuition, expressing caution over entrusting machines with critical life-and-death decisions due to the severe implications of possible errors. The interviews echoed this sentiment, where a solid imperative was voiced for preserving ultimate decision-making authority with military personnel. Trust in commanders was a recurring motif among professionals, pointing to a fundamental belief in the unique capacity of human judgement and intuition to navigate the complexities of combat. Together, these insights from both studies articulate a collective wariness towards algorithmic rationale in dynamic situations and an endorsement of the irreplaceable value of human discretion.

Incorporating thoughts on the ‘responsibility’ node articulates a nuanced understanding of accountability in the era of unmanned warfare. The focus groups crystallised the notion that responsibility for the actions of AWS squarely falls on the individuals within the military hierarchy who authorise their use. This perspective was mirrored and expanded upon in the interviews with military personnel, which fostered a discourse around the locus of responsibility in a landscape marked by increasing autonomy.

Here, the consensus gave extra weight to the belief that the commander who gives the order for the engagement of AWS is the ultimate bearer of respon-

sibility, unequivocally consolidating the human element at the core of accountability. A shared emphasis on the responsibility of the commander and not the system engineer, manufacturer, state or any other actor across both dialogues reflects a collective apprehension about the person in charge.

The discussion also coalesces around the contentious debate of banning versus regulating autonomous weapons. The consensus reflects apprehension about the prospect of pre-emptive measures to curtail such systems. The participants from the focus groups implicitly recognise the perils of unchecked technological proliferation in warfare, advocating a prudent and restrained approach to adoption. Experts, likewise, highlight the imperative for imposing boundaries not simply as a form of regulation but as a safeguard against the dilution of human control in military decision-making and to maintain human agency in warfare.

Finally, these discernible clusters of terms and figures of speech that make up interpretative repertoires illustrate how the participants position themselves and others in narratives and reinforce or challenge societal norms and values (Wetherell & Potter, 1992). By examining these repertoires, I have also relied on analytical discourse to supplement the visualisation map of clusters to unmask and delineate taken-for-granted understandings, trying to transform them into objects of discussion and criticism and, thus, open to change (Jørgensen & Phillips, 2002).

3.4.1 Unity in Uncertainty: Bridging Perspectives on Unmanned Warfare

The three converged nodes – decision, responsible, and develop – align closely with my primary objectives to research perceptions around the use of AWS. They directly address the core questions I elaborated for both studies to explore whether the opinions of professionals about military AI applications could be at variance with those of ordinary citizens. Such a dissimilarity turned out to be not the case. The triad of themes encapsulated the common challenges posed by delegating inherently human tasks to machines and the pressing need to navigate development responsibly. The attribution of characteristics to these themes highlights the tension between human discernment and the recognition of the rapidly expanding abilities of machines.

The confluence in the given discourse underlines the agreement on the inherent dangers posed by autonomous weapons, prompting a resolute push for legal frameworks that emphasise human judgement over AI-powered prowess. The convergence of the themes underscores a shared commitment to preserving human values amidst swift advancements in autonomy. Collectively, these perspectives reveal a uniformity in concerns across lay and expert domains, demonstrating a commonality in the desire to ensure human-centred governance in developing and using autonomous weapons.

Moreover, the unease over military AI unifying ordinary citizens and experts, reflects deep-rooted values and anxieties about technology's role in life-and-death

scenarios. From the perspective of military sociology, this reveals several vital points. The shared concerns highlight a preference for human decision-making, suggesting a resistance to allowing weapons to dictate critical ethical outcomes. The emphasis on responsibility is indicative of a fundamental need for maintaining a regulative framework where actions can be traced to accountable military personnel. AWS challenges accountability, which is central to the functioning of societies to maintain the rule of law. Autonomous systems, with their perceived lack of discernment, bring the sanctity of human life to the forefront, prompting a societal re-evaluation of what is deemed acceptable in warfare.

Furthermore, distrust in AI-driven systems is compounded by the high stakes involved in military applications, where the consequences of error are dire. The introduction of AWS disrupts traditional narratives of heroism, sacrifice, and valour, which soldiers have historically personified. In both studies, a patriotic mentality to fight for freedom was underlined as inconceivable for an autonomous system, meaning that unmanned platforms can hardly ever be equipped with patriotic algorithms.

Given that universal values are widely shared, consensus might be expected to influence attitudes towards not entrusting human lives solely to machines. Even so, attention must be given to the point that experts (**Study III**) may not necessarily possess superior knowledge of AWS-related aspects compared to ordinary citizens (**Study II**). Nonetheless, it is also crucial to acknowledge that the data originates from 2021, preceding the full-scale war in Europe.

However, considering the swift integration of military AI and a new arms race being underway as of 2024 (Hirsh, 2024), it can be assumed that today's information landscape may diverge significantly among these groups. The revealed common ground in the closing remarks encapsulates a dilemma between embracing AI progress and advocating for caution. It suggests a collective consciousness that seeks to balance innovation with ethical considerations, preserving core human values in the face of rapid technological advancement.

In summary, the convergence in the discourses of ordinary citizens and experts may stem from a lack of differentiation in their perceptions and discussions, attributed to the phenomenon's relatively new and complex nature. As AWS continues to evolve and its implications become more evident, we may observe a more significant differentiation and divergence in the perceptions and influences on the discursive repertoires of various groups.

4. DISCUSSION

In the preceding chapters of this dissertation, I have explored the multifaceted dimensions of AI-driven warfare, explicitly focusing on the anatomy of autonomous systems. Through a detailed examination, I have uncovered the intricacies and challenges in defining AI weaponisation and autonomy in these systems, delved into the philosophical and technical aspects of machine autonomy, and scrutinised the complex ethical, legal, and sociological implications they present. The research has revealed an intricate human-machine analogy and a rich tapestry of interdisciplinary perspectives. As I transition into this discussion chapter, I aim to synthesise these findings from an Estonian military perspective¹ and critically examine the future trajectory of autonomous military systems in the context of AI-driven warfare.

Before I move on to the main discussion, I remind you that weapons systems have not yet realised their full potential, in autonomy in particular, but their eventual emergence could represent a pivotal moment in the military field. When such systems materialise, I expect their impact on society to be more profound and far-reaching than perhaps in any other domain of AI application, primarily because the original source of lethal force might no longer derive from a human.

An actual global arms race to produce an AI-powered military is virtually inevitable, and the endpoint of this trajectory is obvious: autonomous weapons will become the *Kalashnikovs* (the most used assault rifle) of tomorrow (Tegmark, 2017). AWS have also been described, with good reason, as the ‘third revolution in warfare’ after gunpowder and nuclear bombs (Russell, 2019). In favour of the inevitability of AI-driven warfare, it speaks to the fact that wars generate massive amounts of data, and it seems tempting to outsource artificial expertise and algorithms that can uncover hidden insights to form a calculated judgement of odds in the dynamic battlespace (Schwarz, 2021).

Hence, if the involvement of AI in military operations appears inevitable, it becomes crucial to ensure that people’s opinions are given due consideration. This is mainly a case where progress must stem from the clash between potential pragmatic benefits and profound moral implications (Purves et al., 2015; Zajac, 2023). The juxtaposition of different arguments is central to ongoing discussions about ethical values, regulations, and the future role of AI in warfare (Bode et al., 2023).

Societies are severely limited in anticipating and mitigating the unintended effects of rapidly emerging technologies without the engagement of good-faith,

¹ The data example from Estonia is relevant because the Estonian Defence Forces (EDF) has deployed and evaluated unmanned vehicles in diverse configurations, positioning Estonia as one of the pioneering European nations to utilise unmanned systems in international military cooperation (Study III). Furthermore, research in Estonia aligns with insights from WEF participants in Davos (Study II), who echoed sentiments similar to those of the Estonian focus groups, which explains why the Estonian empirical evidence is relevant and allows for certain generalisations.

which risks creating blind spots to consequences that will often not be obvious to elites like tech leaders and policymakers (Scheufele et al., 2023). Similar apprehensions were observed in the US; there are genuine concerns among the public about the consequences of a military AI, particularly the fear that humans will have less agency over life-and-death decisions in war (Morgan et al., 2020).

The assertion that inclusive, earnest insights from diverse opinions are crucial is not just theoretical; navigating the uncharted waters of the unintended consequences of weaponised AI is necessary to gain trust and foster greater confidence and acceptance. In the ensuing subsections, I intend to stir a vigorous debate by bringing military personnel into the discourse and including ordinary citizens, typically side-lined from important facets of lethal force, to engage in the dialogue to represent the settled way of thinking of ordinary people and end-users. This approach is about anticipating and mitigating impacts and fundamentally rethinking how emerging technologies might reshape the rules of *ius in bello*, ethical boundaries, and societal norms. This chapter's goal is to confront the uncomfortable questions and hidden assumptions that lurk beneath the surface of autonomous weapons.

4.1 Decisions of AI-driven Autonomous Systems in the Military Domain

Let us consider the implications of the top three aspects identified in the literature review (Figure 1): compliance with IHL principles, responsibility, and life-death decision-making. At the heart of these concerns lies a fundamental question: are we comfortable allowing machines to make decisions once solely the remit of humans, especially in contexts governed by the laws of war? First, when discussing machines complying with IHL principles, we are delving into an area steeped in human judgement. These principles were conceived with human understanding and discretion in mind.

From the very start of armed conflict, there is an interplay of possibilities, probabilities, luck, and the evil of human nature (Clausewitz, 2004), making war logically irresolvable using computing power and datasets alone. I argue that programming a machine to interpret such principles is not merely a technical challenge but an ethical one because ethical norms are intricate, often subjective, and differ across cultures about which people constantly argue. How then do we encode the nuances of human ethics into binary code? This question is not just academic; it has existential implications for enforcing the core principles of IHL and protecting non-combatants in the age of autonomous weapons to come.

Already in 1976, Weizenbaum highlighted a crucial distinction between decision and judgement for a computer. According to Weizenbaum, deciding is a computational activity that can be programmed, essentially a quantitative process, and does not have the basis for forming values. However, the ability to choose is a product of human judgement guided by values, not mere calculation, which makes us human. Machines that lack cognitive qualities like compassion

and wisdom, should not be entrusted with making important decisions (Weizenbaum, 1976).

The discussions from the focus groups (**Study II**) revealed a similar nuanced understanding inherent to judgements, emphasising that the efficacy of a decision – whether made by humans or machines – varies greatly depending on the context. A critical issue raised was the current inability of machines to replicate the ethical judgement and emotional intelligence innate to humans. The discussions underscored the immense challenges in developing autonomous systems that parallel human intuition and moral reasoning, particularly in high-stakes military operations.

However, the genuine concern is whether machines have inadvertently or intentionally been granted authority over matters where they lack expertise. The insights gathered from experts (**Study III**) discuss the irreplaceable role of human judgement, underscoring the risks associated with delegating grave decisions to machines in situations where human limitations like stress and fatigue play a role. There is a strong consensus for a cautious approach to the impact of AI on military command structures, pointing out the nuanced and context-specific nature of decisions in dynamic warfare that machines cannot yet replicate. Ultimately, the discussions converge on the view that while autonomous systems can support, they should not transform or replace the essence of commanders' decision-making in warfare (Lee et al., 2023). That military perspective reflects a deep-seated conviction about human leadership's moral responsibility and agency, ensuring that the final say remains firmly in the hands of an officer.

4.1.1 'Knowledge' – Awareness that a Consequence will Occur

In human decision-making, the mental element of accountability (Cassese et al., n.d.) is clear if the material elements are committed with intent, although it is not always straightforward. Those lines become blurred, however, when machines are introduced into the equation. If a machine makes a decision that leads to the loss of life, who is responsible? The programmer, manufacturer, operator or the machine itself? The prominence of responsibility, in my view, signals an urgent call to clarify these lines of accountability before we reach a point of no return.

The intuition from ordinary citizens (**Study II**) concerning legal liability identified the owner of the weapons system as bearing primary accountability for the actions of these systems rather than the developers or manufacturers. Military leaders were pinpointed as key figures who instruct the use of autonomous weapons and should be accountable for the outcomes. This viewpoint underscores the challenge of accountability when decisions are distributed across a hierarchy, leaving those at the bottom – the executors – often unfairly condemned despite merely following orders. Rather, the participants implied that responsibility is collective and distributed among all in the chain of command, including those who set the broader strategic goals. Furthermore, the prevailing logic of the studies implies that a system's capability is insufficient if a machine cannot

encompass the emotional and cognitive elements of decision-making by stressing that the responsibility is qualitative and deeply human.

It has been firmly contended that, regardless of the artificial capacity to make ‘correct’ decisions, machines should not be entrusted with certain responsibilities (Weizenbaum, 1976). This viewpoint was expressed much earlier and is particularly pertinent today as AI is increasingly integrated into the use of lethal force and targeting roles. The crux of the issue is not the technology *per se* but rather the ethical choices that necessitate its use. These decisions require value-based judgements, intuition, and wisdom – attributes that Weizenbaum contends are outside the scope of computer simulation (Weizenbaum, 1976).

The interviews with military personnel (**Study III**) delved into the context of distinction and targeting in military operations. The study indicates an understanding within the Estonian Defence Forces (EDF) that the act of decision-making, even when delegated to an autonomous system, does not absolve the human operator of responsibility; in other words, the person behind the weapon, despite its autonomy, remains accountable. The officers grappled with the idea of machines replicating human decision-making, which is not merely logical or based on binary data but is informed by an interplay of factors. The study indicated scepticism about where machines may potentially apply principles of IHL and signalled a significant apprehension about the realm of distinction and targeting that cannot be transferred to machines and must ultimately reside with humans.

The knowledge needed to be aware of consequences is inherent to decisions in combat and is a critical term in jurisdiction over persons for the most severe crimes of international concern (Cassese et al., n.d.). More precisely, awareness subsumes the mental element necessary to constitute an offence if the material elements are committed with intent and knowledge (Colorio, 2021). However, in terms of employing AWS and the hotly debated accountability gap (Verdiesen et al., 2021), I posit that a theoretical scenario exists in which charges against a military commander should be dismissed if the individual lacks intent, knowledge, or the ability to control the outcomes of a given situation. Therefore, the accountability gap can be closed because the algorithmic use of force is hazardous, and it does not matter if a person neither intended nor anticipated unlawful consequences. By saying that, I agree that the outcome was also accepted when the decision was made to accept the high risk associated with using autonomous weapons (Leveringhaus, 2021).

The military is not the only one grappling with responsibility issues, risking someone’s life. A study from the medical field explores how concerns about legal liability influence a physician’s decision to use or not use AI in clinical practice. The study found that physicians might use AI more in ‘low uncertainty’ cases, where they are already confident in their treatment plan, and avoid using it in ‘higher uncertainty’ cases due to the risk of legal repercussions if they deviate from the AI recommendation (Dai & Singh, 2023). The study recognises AI’s potential to improve healthcare, yet the current legal framework discourages its optimal use, particularly in sensitive cases where AI precision could be most beneficial.

I think precaution forces a stalemate driven by the desire to mitigate legal risks rather than the AI guidance. A cautious approach will be adopted in placing trust in AI's judgement to sidestep potential legal complications. The military leader may choose not to use AI or follow its recommendations in straightforward cases because of the fear of malpractice lawsuits rather than because AI has provided incorrect advice. My concern is that such caution might impede the full potential and advancement of AI assistance in tough decisions because obscurity in accountability limits the use of AI to avoid legal consequences rather than exploring its capabilities to enhance decision-making in more harsh cases. I do not dare to voice solutions, but that stalemate could lead to a slower rate of innovation and integration of AI into clinical and military operational practice.

To my knowledge, no comprehensive study currently investigates the frequency and impact of mistakes made by military officers in interpreting AI outputs, nor how often they alter their decisions after reviewing AI recommendations. Given the increasing reliance on AI in these critical fields, such a study would be precious. It would shed light on the reliability and influence of AI in high-stakes decision-making and also guide the development of protocols for AI-assisted decisions in military settings, akin to those in healthcare. The findings could significantly minimise the potential risks associated with AI reliance and enhance decision-making processes in these domains.

4.1.2 The Pivotal Moment: Moral Judgements in Crucial Scenarios

The frequency and cross-referencing of life-death decisions appearing in my thesis cannot be overstated. Here, the stakes are the highest imaginable. Entrusting machines with the power to make decisions that could end a human life brings us face-to-face with stark differences in decision-making. Humans, though not necessarily every human, can weigh the moral and emotional consequences of their decisions; machines operate within the parameters we set for them, no more, no less.² However, if it is not possible to resolve moral disagreements by rational argumentation, we should understand what causes our disagreements, what we have in common and on what basis we can build consensus or make compromises (Sutrop, 2020).

Each focus group (**Study II**) had participants trying to argue that autonomous systems can potentially remove soldiers from direct harm and, conversely, that algorithmic warfare leads to more deaths. These contrasting views were rich with nuances as integrating AI into the targeting cycle has sparked reactions shaped by hope, scepticism, and ethical unease. For instance, focus group members viewed AWS as a pathway to reduce civilian casualties and potentially minimise

² A note: machine learning models, as they exist today, do not possess independent thought, consciousness, or the ability to understand moral consequences mirroring cognitive abilities as humans do. The ultimate question has been circled for a while: can machines ever earn trust to understand the value of human life? From my perspective, it is an unbalanced question because too many people cannot reach that trust, and many never will.

collateral damage. Others, however, were deeply troubled by the prospect of algorithms, fearing a depersonalisation of warfare and the loss of human judgement. This dichotomy reflects a broader societal grappling with the essence of war and what it means when machines take an active role on the battlefield.

The transition from human to machine in warfare raises substantial concerns about people's trust. Trust in military officers has traditionally been rooted in the belief in a commander's competence and integrity (Yeşilbaş & Çetin, 2019). Introducing AI disrupts this foundation, challenging the understanding of war, which has long been a profoundly human affair. With AI, ordinary citizens (**Study II**) experience an underlying anxiety that the abstractness of algorithms could obscure the tangible human costs of conflict, leading to a disconnect from the real.

Several participants were convinced that algorithms are abstract and rely on data and calculations, which do not encapsulate the human experience of suffering. For instance, there was a fear in many groups that decision-making might become too detached, focusing on efficiency or outcomes as computed by machines without appreciating the human lives affected by these decisions. Moreover, the opacity of AI systems complicates efforts to foster transparency and accountability, further eroding public trust.

As AI continues to advance, a critical sociological task will be to navigate these perceptions and develop narratives that address these valid concerns. Ensuring robust ethical standards, transparency in AI deployment, and ongoing public engagement are crucial steps towards maintaining a societal understanding of war that aligns with our collective values and the principles of human dignity. The challenge in AI-driven judgement lies in programming systems that interpret complex social interactions and make ethical decisions with often incomplete or ambiguous data. That is crucial for reducing mistakes and unintended outcomes in military operations to effectively navigate the subtleties of human social reasoning and ethics. Therefore, the dialogue about AI in military use must be as much about preserving human oversight and ethical integrity as it is about technological progress.

Trust plays a pivotal role in human interaction with technology (Sztompka, 1999). The unfamiliarity or lack of understanding about autonomous systems often leads to hesitation or outright mistrust. This scepticism towards the unknown is a natural human reaction. I mean that simply introducing AI in warfare without proper education or acclimation will not suffice to mitigate individuals' inherent distrust. A concerted effort must familiarise people with the benefits of autonomous systems and safety measures, ensuring a smoother integration of AI technologies in military operations and beyond.

Trust in military AI should be calibrated by enhancing confidence among individuals. That can be achieved by regularly disseminating relevant information, fostering open communication about potential shortcomings, and maintaining transparency in the input-output ratio. Such measures are critical in building a robust understanding and acceptance of AI's role in military operations, emphasising the importance of confidence as a cornerstone of trust in AI systems.

Moreover, the military leadership should pay heed to people's opinions and beliefs about algorithmic warfare, not the risks of maintaining the moral centre of gravity in utilising autonomous systems.

The notion of determining life-and-death outcomes introduces discomfort. While autonomous systems can aid in outlining potential actions, the ultimate decision-making power regarding human lives should not be vested in them (**Study III**). The experts do not see ethical issues with the systems *per se* but imply that the ethics lie in the use and control by the decision-maker, recognising the potential for misuse. The interviewees pondered the desensitisation to killing that autonomous ground systems might induce and expressed an unease at the thought that war could become less 'real' as has already happened with unmanned aerial systems (Strawser, 2013).

It follows that, in the realm of unmanned warfare, the act of remotely executing lethal force may inadvertently ease the psychological weight associated with the act of taking life and disregard the gravity of a kill. However, empirical research indicates that the mere physical distance from a target does not significantly alter an individual's propensity to engage in lethal action. The effect of distance on the willingness to kill was statistically insignificant, suggesting other factors might play more critical roles in such behaviours (Rutchick et al., 2017).³ In this case, moral subjectivity may hold divergent views on what constitutes ethical use of AI in military contexts.

It has been argued that sometimes people choose the same option but provide different justifications from various perspectives (Sutrop, 2020). Some might base their choice on utilitarian considerations, focusing on the benefits for others, while others might justify their decision on moral grounds, emphasising the importance of doing what they believe to be inherently right. The variety of values or principles that can guide individuals' choices highlights the complexity of simulating decision-making processes artificially.

The lethal quality of AI should be alarming; societies have a finite window of opportunity to find ways to engage in good-faith debates and work towards meaningful AI to ensure these challenges do not overwhelm them (Scheufele et al., 2023). In synthesising these views, the overarching implication is that while autonomous systems can revolutionise military affairs, distancing soldiers from direct participation in hostilities, there is an implicit call for a balance where technological innovation must be matched with moral responsibility and human oversight.

³ An experimental setup explores the psychological effects of technology-mediated remoteness on individuals' willingness to engage in an ostensible ladybug-killing task. Although participants killed more insects when they were in a different room than their targets, the study found that the isolated effect of physical distance did not influence behaviour as anticipated. The failure to observe a significant effect when ostensibly locating the targets in another state suggests that it is not physical proximity alone that drives the observed effects. Instead, the study posits that the effect of being in a place different from the targets might be more attributable to psychological remoteness rather than spatial distance.

4.1.2. A 'Black Box' – Input and Output of the Mind and Algorithm

Our brains possess a remarkable faculty known as *ad hoc category construction* – a concept highlighted by Lisa F. Barret (2021) in an influential MIT Technology Review article. She argues that cognitive ability allows the brain to rapidly utilise past experiences to form abstract categories. For instance, categorising a feather as a symbol of honour is based not on its physical attributes but on its functional and symbolic significance in specific contexts. However, unlike the human brain, current computational systems have yet to master this. Machines are incapable of recognising abstract concepts, such as a feather symbolising honour, as this understanding transcends the object's inherent properties and lies in the realm of abstract categorisation, a uniquely human cognitive process (Barrett, 2021).

Although machines cannot recognise abstract concepts, their application in military contexts shows significant advancements. AI-driven systems update their understanding of the battlefield in real time, making more informed decisions by integrating and processing data at unprecedented speeds. This rapid processing allows for more informed decision-making for commanders, underscoring the significance of data integration and interoperability in modern warfare. I should emphasise that such software plays a crucial role as a connector, bridging interoperability gaps and enabling coordinated action among multinational forces. Furthermore, AI programs are in a state of perpetual evolution, continuously improving through model refinements and the ingestion of more and better-quality data. This ongoing learning process means that algorithmic analysis can potentially identify connections in data about enemy forces that would be imperceptible to human analysts.

From my perspective, it is vital to explore the intricate landscape of rationale that guides us in reaching an informed conclusion. A fascinating comparison emerges between the human mind and the structured realm of convolutional neural networks (CNN),⁴ suggesting that human observers and CNNs prioritise more similar features during object recognition of animate objects (van Dyck et al., 2021). This statement points to the promising prospects of merging neuroscience with AI technologies, suggesting that such an alliance could usher in revolutionary advancements in understanding human cognition and enhancing AI's mimicry of neural processes. Researchers are hopeful that by studying and developing CNNs inspired by neural processes, they might gain valuable insights into the brain's complex mechanisms, an area still not fully understood (Xu & Vaziri-Pashkam, 2021).

The gist of this is to address the opacity of AI's decision-making processes to the internal, unobservable thought processes of soldiers in combat to understand the cognitive mechanisms underlying human decisions and the algorithmic pathways leading to AI-generated conclusions. I advocate enhancing the transpa-

⁴ CNN stands for a type of AI designed to recognise patterns in images inspired by the human visual cortex and designed to learn features from images adaptively. CNNs excel at tasks such as object detection and facial recognition and aid autonomous vehicles in navigating by interpreting the visual data from their surroundings (Yamashita et al., 2018).

rency behind decisions by AI-driven systems capable of explaining their reasoning because it facilitates bridging the accountability gap and fosters greater trust in AI's role in military operations by demystifying its judgement process.

The human mind, known for its abstract reasoning, emotional depth, and adaptability, is in stark contrast with the data-focused and often rigid nature of algorithmic rationality. Despite a profound and unmistakable disparity between these two, both embody unique approaches to What if we were to compare an autonomous weapon to a soldier with a rifle? In this parallel, both can be lethal in achieving military objectives, and there is an extent of human oversight under a chain of command expected to operate within the confines of the laws of war. True, the human mind is not algorithmic; that is, having a set of rules to decide, but in both cases, there is a degree of autonomy.

Keep in mind that soldiers are those who still commit atrocities. The '*black box*' in soldiers' minds can bring us back to where we started with *ius in bello*. We are concerned about whether autonomous systems can adhere to IHL principles, but in the 21st century, we need only one armed conflict to prove how easily humans can fall back morally to the Stone Age by committing war crimes, as evidenced by Russia's war in Ukraine. Now that tech giants have turned Ukraine into an AI war lab (Bergengruen, 2024), this dissertation aims to provide acute awareness about autonomy in weapons and add pragmatic value to research in military AI.

Consequently, being acquainted with juxtaposing, soldiers may not always make their plans or thoughts known, leading to unpredictable behaviour. Similarly, how CNN operates cannot be immediately apparent. Despite the uncertainties, a parallel emerges between a soldier's actions and the CNN learning methods, such as backpropagation.⁵ To connect backpropagation to the human visual cortex, let us consider an example of learning and adjusting based on errors or feedback.

Imagine a soldier learning to differentiate combatants from non-combatants. When a soldier mistakenly perceives a child holding a rifle as directly participating in hostilities, a superior steps in to correct the misunderstanding. The clarification includes the possibility that the child might have simply found the rifle by chance, not intending to participate in the conflict. The visual cortex assimilates this feedback, refining its perception to differentiate between mere possession of a weapon and active engagement in military operations. Over time, through case studies and correction, the soldier becomes more adept at recognising these situations.

Now, this learning process is somewhat mirrored through backpropagation. When the network identifies an unlawful target during training, the error is calculated using a loss function. Backpropagation takes this error and propagates

⁵ The backpropagation computes the gradient of an objective function with respect to the weights of a multilayer stack of modules. The key insight is that the derivative gradient of the objective with respect to the input of a module can be computed by working backwards from the gradient with respect to the output of that module (LeCun et al., 2015).

it back through the network, adjusting the weights of the connections between the neurons (akin to adjusting the soldier's understanding in the example above). This process gradually improves the network's ability to identify military targets.

So, there is potential for AI trained in warfare to become more attuned to human nuances by learning from the intersection of the experiences of soldiers and the capabilities of CNNs through backpropagation. This approach could leverage AI's ability to learn from mistakes, potentially reducing human errors influenced by the psychological impacts of war. By integrating insights from human cognitive processes and AI's analytical precision, it is possible to enhance decision-making in combat scenarios, minimising the risks associated with stress-induced miscalculations. While the comparison between human cognition and AI algorithms is disputable (Xie, 2023), it merely highlights the potential of autonomously exerted force in unmanned warfare.

Another aspect that concerns me is the algorithmic 'brain' lifecycle. Understanding the lifecycle of an AI model or system is essential for the needs of the end-user and the operational context, as data quality directly impacts the model's performance. Once deployed, continuous monitoring is essential to ensure consistent performance, with periodic updates and retraining necessary to address issues like model drift (i.e., when a statistical property of the target variable, which the model is trying to predict, changes over time). This lifecycle approach is exceptionally critical in autonomous systems and must ensure that military AI solutions are robust and ethical and deliver sustained value.

4.1.4 A Human-Machine Teaming

Striking the right balance between human and AI competencies is a crucial future consideration. While AI can enhance capabilities, there is a belief that greater autonomy does not necessarily equate to more capability, but human-machine teaming is considered to deliver the best outcomes in terms of effectiveness and resource use and is the preferred approach to AI adoption (Defence AI Strategy, 2022). It is argued that humans do not 'collaborate' with a toaster to make toast, yet the field of collaborative AI considers instead the term 'collaboration' to be self-evident (Evans et al., 2023).

The 'collaboration' and 'teaming' in human-AI interactions are nuanced, hinging on the semantics and perceived roles of humans and AI in these interactions. On the one hand, 'collaboration' implies a partnership where parties work together, share responsibilities and contribute equally to the outcome. In a collaborative scenario, the AI is not just a tool but a partner that brings its unique capabilities to the table, complementing human skills and insights. The absence of a designated leader suggests a more fluid and dynamic interaction, where leadership can shift based on the project's task or phase, aligning with the idea that AI can operate autonomously, making decisions or offering insights without human oversight at every step.

I share the 'toaster' irony and prefer the term 'teaming'. In this sense, humans are typically in the leadership role, guiding the task and making final decisions.

This distinction is crucial as it acknowledges the current limitations of AI and emphasises the importance of human oversight, recognising the unique contributions of both while acknowledging their distinct roles and limitations. I am unsure how this teamwork plays out, but in principle, teaming allows for an interplay to compensate for the other’s limitations.

It is even argued that the thin line between humans and machines evaporates and will vanish soon (Moelker & Schenk, 2018). I would not put that so resolutely, although I do not exclude the possibility of hybrid entities. I see that the human role will not be diminished but somewhat enhanced, as machines can free up human cognitive and creative resources by taking over more routine and data-intensive tasks. By working in tandem, we can ensure that technology is used ethically and responsibly, aligned with our values and societal goals. It sounds trivial, but it will open new avenues for exploration and discovery, which would be unattainable by humans or machines alone.

An experiment demonstrated that pathologists, when assisted by AI, show a notable increase in diagnostic accuracy in prostate biopsy assessments, highlighting that beliefs about AI’s utility may play a more decisive role (Meyer et al., 2022). This experiment emphasised that pathologists made faster decisions with AI support. These findings underscore the growing importance of AI in enhancing medical diagnostics while also urging me to consider how military professionals communicate AI capabilities. One method could be employing a case study to examine AI-powered weaponised platforms considering the following scenario I have devised: integrating AI into military decision-making, especially regarding balancing technological inputs with human judgement.

So, let me construct a hypothetical scenario: a coalition of states has enhanced its presence in the area of operations. The multinational battlegroup has adopted a new system that enables the management of AI-infused decisions. The system is manufactured and pre-trained in country A. It has undergone appropriate testing verification and has been validated, although training data is restricted and has not yet been applied in combat. No ROE specifies when and how computer-aided systems can identify and engage targets, ensuring that actions are legally and ethically sound and in line with the objectives and policies of the military operation. As it was deployed in armed conflict, a target identification software spotted a high-value target and reported it to HQ. The technical specifications are as follows in Table 6.

Table 6. IHL core principles adapted to the case study.

Distinction	The precision of targeting is at 75%, which aligns with technical standards for positive identification set by country A’s regulations.
Proportionality	The system assessed the extent of injuries and collateral damage, forecasting a 75% likelihood of compliance with the proportionality standards mandated by country A.
Military necessity	The system advised using an air-dropped bomb to destroy the target and achieve the objective.

The case study continues. Recce squad from country B is also on the scene (incl. targeteers) and copies the targeting information. The squad leader reports to battlegroup HQ that from their estimation, identification is 50% and suggests that an air-dropped bomb is unnecessary because it might not affect the target and that there is collateral damage potential to friendly forces and civilians in the vicinity. Considering this, what information should the multinational unit commander from country C consider or trust more before engaging the target?

Case study No 1: the same scenario, but the estimations are vice versa. The recce squad members are highly confident about identifying the military target at 75%. According to their estimation, there is no risk of proportionality and collateral damage. An air-dropped bomb is appropriate, and the consequences of not attacking the target do not accomplish the military objective. According to calculations by the targeting software, identification is 50%; weaponeering does not consider the desired effects against the target and poses real collateral damage to friendly forces that is not justified by military necessity. What information should the commander consider now or trust more before engaging the target?

In exploring further complexities of AI-human interaction, two additional sub-cases expand on the initial scenario to assess adaptability and trust in AI systems. In one scenario, HQ directly shares the AI-generated target summary with the unit, leading to varied responses – some units may show tentative agreement with reservations about AI's accuracy, while others may express strong disagreement, favouring human judgment over AI. This variation probes how differing levels of agreement with AI influence command decisions and trust in AI recommendations. Another scenario sees HQ sharing the unit's assessment with the AI, prompting the AI to adjust its advice by incorporating human feedback, mitigating risk levels without recommending engagement or modifying its parameters to reduce potential errors. These sub-cases challenge commanders to reconsider their strategies based on an AI that dynamically integrates human assessments, exploring how such integration affects operational decisions and the overall effectiveness of AI support in military environments.

These case studies intend to illustrate a scenario of human-machine teaming and the complexities of integrating AI into military decision-making, especially regarding balancing technological inputs with human judgement. If I put myself in that commander's position, I do not have a firm answer as to what I would do or trust more. I suspect I am inclined to trust my unit rather than the machine's assessment. However, it is also true that a commander should not exclusively rely on either human judgement or AI but should consider combining both case by case. In line with what has been said, it drives me to theorise that relying on AI-enabled autonomous technologies depends on many variables, so it is always contextual to summarise very briefly.

While acknowledging that my perspective may not seem immediately groundbreaking, the thesis emphasises that the pivotal factor in employing autonomous weapons lies in the human intention behind their use. Moreover, disregarding military AI's entrenched role and ongoing evolution would be a fundamental error. I cannot think of ways to regulate algorithms or particular deep learning

networks;⁶ it is just not realistic, but it is feasible to regulate specific applications that violate the laws of war or are against the public conscience derived from the Martens Clause (Hensel, 2016).

Study III highlights that while orders and goals are given to military personnel, the exact methods of achieving them are not always dictated. This approach is based on the belief that the training and education provided to officers enable them to make independent decisions and act accordingly, which motivates individuals. The attitude suggests trust in human judgement and capability and a recognition of the unique value that human factors bring to problem-solving and task execution instead of the mechanical and predetermined paths of machines.

Reflecting on this discussion, it appears that the essence of integrating AI with human capabilities lies not in pitting humans against machines but in combining their unique strengths. AI's ability to model ideas and simulate behaviour complements human intuition and creativity. By integrating the best aspects of human sensibility with the computational power of AI, we can create a synergistic relationship in the military domain. This fusion enhances our capacity to solve complex problems, make more informed decisions, and innovate at a pace far beyond what either humans or machines could achieve independently. The future, therefore, should focus on leveraging this teaming, maximising the potential of both human intellect and AI efficiency.

To summarise this subchapter, human-machine teaming is not just a practical approach but a visionary one. It embodies a future where technology augments human potential rather than replacing it, paving the way for advancements that are both profound and beneficial to the military. I recognise the necessity for accelerating human-aware AI (Sourati & Evans, 2023), enabling the system to comprehend the nature and purpose behind its actions in pursuing human objectives.

In conclusion of the discussion chapter, these three aspects – compliance with IHL principles, responsibility, and life-death decisions – do indeed point to a broader theme of concern: the tension between machine efficiency and human value judgements. This chapter aimed to delve into the realms of autonomous weapons by using an interdisciplinary observation of the intricate dynamics of human vs. machine decision-making, the quest for accountability in autonomous actions, and the vast sphere for the stringent regulation of the unknown. As we stand on this technological precipice, it is crucial to interrogate not just the capabilities of machines but the values we as a society want to uphold. It is not a question of whether machines can make decisions but which ones and under what moral framework we will allow them to operate.

⁶ Deep learning is a subset of machine learning, which attempts to mimic the workings of the human brain's neural networks to process data and recognise patterns for decision-making. Deep learning models are composed of many interconnected layers (also called neurons), each layer transforming the data in a hierarchical manner. These layers allow the algorithm to learn increasingly abstract features from the input data (Sarker, 2021).

CONCLUSIONS

Undoubtedly, given the potential of AI's progress, autonomy built in autonomous systems revolutionises how armed forces will operate on the battlefield. Autonomous weapons speak to the fact that the military domain is ground zero regarding AI applications because it directly concerns life-and-death decisions outside human agency, making it a focal point of ethical debates and legal discussions. The indication of a shift in the nature of combat is demonstrated by the transition from human-intensive warfare to one that heavily involves drones, robots, and AI systems.

The armed forces are gradually reducing the human presence on the battlefield and relying more on autonomous systems, which conduct repetitive or prolonged tasks without fatigue, allowing for sustained operations beyond the physiological limits of humans. However, removing soldiers from direct combat increases the risk of lowering the threshold for unmanned warfare. The reliance on technology also raises its vulnerability, caused by malfunctions, hacking, or cyber-attacks that might disable or repurpose them.

After delving deep into the multifaceted dimensions of autonomy in weapons, it becomes evident that military AI is revolutionising defence capabilities, and moving closer to unmanned warfare. As the arms race in AI weaponisation intensifies, the armed forces are encountering pressure to incorporate autonomous technologies into combined operations across various domains. The increasingly tangible reality of AWS predicts that human-machine teaming will become the focal point of the use of force in the military domain. I will now consolidate my findings and insights by answering the research questions presented in the introduction.

RQ1: What are the main aspects associated with introducing autonomous weapons, as identified in the systematic review? (Study I)

The following aspects underpin the apprehensions surrounding the integration of autonomous weapons into warfare in significant order: compliance with IHL principles (distinction, proportionality, military necessity, humanity), accountability gap, a life-and-death matter, meaningful human control, human dignity, arms race, unpredictability, and the threshold of armed conflict.

a. What is the argumentation underlying these aspects? (Study I)

The underlying argumentation is that autonomous systems either can or cannot adhere better to IHL principles than humans. First, the computations required to apply IHL principles exceed the programming and learning capabilities of algorithms. On the contrary, for better rule adherence, algorithms can process sensory information without being influenced by emotions or human biases that might skew or misrepresent data to align with pre-existing beliefs. Next, almost half of the argumentation is related to the legitimacy of autonomous weapons, giving the impression of impunity, implying that there might not be full accountability for

war crimes. The other half of the debate is divided by normative ethics, among which the deontological position provided the primary foundation for prohibiting autonomy in weapons. Namely, life-and-death decisions outside human control should be reason enough to impart a solid opposition to AI weaponisation.

RQ2: What constitutes autonomous systems, and how is weapon autonomy conceptualised? (Study I–III)

To a great degree, autonomous systems, within this context, should be understood as a military force multiplier that perceives the surrounding environment, makes sense of it to some extent, identifies military targets, and is ultimately capable of making decisions to apply lethal force without direct human intervention. Understanding weapon autonomy requires an examination of the constantly evolving intersection of technology and warfare. At its core, in this thesis, autonomy refers to the capability of a weapon to operate and make decisions without external control or influence, based on perceiving the environment and processing information for pursuing a course of action in alignment with those pre-programmed and supplemented learning criteria.

a. What are the characteristics of decision-making by machines and humans, mainly concerning matters of life or death, mirrored in the views of the study participants? (Study II–III)

Human versus machine decision-making involves intricate considerations, especially in matters of life or death. The study participants believed such decision-making is inherent to humans and that machines lack moral agency and cannot cope with the nuances that situational awareness requires. In this regard, both decisions are likely to be made from different perspectives: human moral values vs. computational calculation. For every participant, such meticulous judgement is always circumstantial, depending on information accuracy, and needs highly context-dependent interpretation. Therefore, it can be concluded that human presence is inevitable when more delicate and ethical decisions must be made.

b. Who, from the perspectives of the study participants, would be potentially responsible if war crimes are committed due to using autonomous weapons, and why is there the possibility of an accountability gap in such cases? (Study II–III)

The accountability gap in the context of autonomous systems refers to the challenge of assigning responsibility for actions driven by algorithms, creating ambiguities about enforcing traditional legal accountability frameworks adequately. The consensus in the studies was that the soldier, machine, or anyone not in command is just a means to execute an order, which means that any person in the military hierarchy who gives direct orders to use an autonomous weapon should be held responsible for the actions resulting from that order. The crux of the participants' concern is ensuring that, in the realm of unmanned warfare, accountability is not side-lined by technological advancements.

c. Why do participants consider the development of autonomous weapons ethically sound or morally unacceptable, and what are the argumentative reasons for moving forward with the authorisation of their development or bringing pressure to bear on a ban? (Study II–III)

Whether developing autonomous weapons was right or wrong was not solely an ethical problem for the participants. Instead, the studies pointed out that development depends very much on circumstances but mainly on the purpose of usage because machine behaviour cannot simply be categorised as ethical or unethical. Overall, banning autonomous weapons was not directly favoured, as this was generally considered unlikely to happen, but putting measures and standards into force to monitor development was the consensus.

d. What dilemmas arise in the era of weaponised AI among ordinary citizens? (Study II)

The participants primarily found autonomous systems preferable to their own troops because they could save the lives of fellow citizens. By contrast, the participants prefer to fight against enemy soldiers than enemy robots mainly because human behaviour is less frightening and more predictable than robots. Yet, these options illustrate how paradoxical preferences can be. Similar systems will most likely be used against both parties and, at some point, humans must confront machines on the battlefield.

RQ3: What, according to the expert assessments, is the projected impact of autonomous systems and the conduct of future military engagements? (Study III)

The experts believe that the machine decision-making process will be very problematic or impossible since even commanders struggle with decisions with limited time and information. The general understanding was that future utilisation of AWS will not necessarily change the art of war. The argument is that warfare is, in the end, a struggle of will, and machines do not have free will; it always remains between human leaders who can eventually impose their will to use force when necessary. Overall, experts think it is still too early to discuss the acquisition of AWS, especially if they cannot replace the main functions of troops and are currently more resource-intensive than anticipated.

Ideally, the introduction of unmanned warfare should streamline military operations and reduce the need for extensive human intervention. However, it seems that, at this stage, the opposite is true: maintaining and operating autonomous weapons demands significant human labour and additional resources. For instance, military personnel must enforce algorithmic decisions or recheck the intentions behind machine-generated actions, contrary to the objective of deploying AWS to reduce human workload and enhance decision-making efficiency and resource allocation. Therefore, it is crucial to reassess military AI implementations to ensure autonomous systems fulfil their intended purpose of resource

optimisation, making unmanned warfare an enabler rather than complicating military operations.

Integrating military AI within forces with a heavy reliance on reservists, as observed in Estonia's situation, it is imperative to carefully consider the trade-offs between allocating funds towards conventional military equipment and investing in more expensive AI systems. The current reality is that soldiers, not weaponised AI systems, are the ones engaged in combat scenarios. Prioritising investments in conventional equipment ensures that soldiers (reservists) are well-prepared and effectively equipped for the realities they face. That does not negate the potential benefits of autonomous technologies but highlights the need for balance. Such an approach would ensure that, in particular, land forces remain versatile, responsive, and capable, leveraging the best of both conventional and emerging technologies to meet the multifaceted challenges of unmanned warfare.

a. What causes uncertainty about the algorithmic rationale in autonomous systems? (Study III)

The uncertainty primarily stems from an algorithm's inability to replicate the nuanced rational ability inherent to human judgement. This limitation in distinguishing combatants becomes critical in urban warfare settings where adversaries may blend with civilians. The indeterminacy surrounding the reliability of algorithms points to a fundamental concern over AI's incapacity to grasp the ambiguous nature of warfare. In addition, while autonomous systems lack human frailties such as fatigue and stress, thereby potentially offering greater military sustainability, the consensus remains that maintaining human oversight over life-and-death decisions reflects a deep-seated conviction of *ius in bello* inherent in warfare.

The study participants have a cautious approach towards autonomy, emphasising the irreplaceable value of human empathy, moral reasoning, and the context-specific nature of military decisions that machines currently cannot replicate. Furthermore, this is far from the only complicated choice to be made, as artificial decision-making can construct dilemmas by default. For instance, an AI-enabled system reports positive target identification, while a manned unit on site does not confirm the same military target. Inconsistent information creates a situation in which deciding which one to trust tests commanders' moral convictions to make a decision.

b. What is the unknown the military is expected to mitigate in unmanned warfare? (Study III)

Incorporating additional technical staff and advisers into military operations to navigate the complexities of autonomous systems appears indispensable for ensuring decisions are both legally sound and ethically justified. However, the practicality of such integration faces constraints due to the limitations on HQ capacity, particularly within the confines of tactical unit sizes. Expanding staff to include legal and ethical advisers, alongside software engineers to manage auto-

nomous systems, could potentially inflate staff numbers to a scale that renders military leadership impractical.

Moreover, maintaining situational awareness and ensuring synchronised movement between manned and unmanned units introduces complexity that may compromise the concept of operation. That underscores the inherent difficulty in achieving effective coordination within units in the dynamic and unpredictable nature of warfare. The indispensable value of human judgement and morale, which autonomous systems are yet to replicate, further complicates the mission statement and integration of unmanned units into task organisation. Training autonomous platforms in military ethics and the ROE represents a formidable challenge, particularly given the diversity of legal caveats that could underpin the operations of a multinational alliance. This diversity raises questions about the feasibility of establishing a universally accepted ethical framework for deploying autonomous systems.

Lastly, anticipating an algorithmic rivalry necessitates the development of AI systems with enhanced predictive properties capable of outmanoeuvring adversary AI. This strategic imperative highlights the need for a sophisticated approach to AI development that prioritises the ability to foresee and counteract enemy threats. The military is expected to reassess the integration of autonomous systems into multi-domain operations, emphasising the need to balance technological advancement with ethical, legal, and sociological considerations in a multinational coalition.

RQ4: What are the converging opinions on autonomous weapons based on the insights of ordinary citizens and experts? (Study II–III)

The analysis of both studies reveals an intersection around the central themes of ‘decision’, ‘responsible’, and ‘develop’ in the context of AWS. Citizens and experts deliberated the shift from human to algorithmic judgement, particularly in using lethal force, which is unacceptable. Another convergent point is why being accountable in unmanned warfare is essential, reflecting a lack of tolerance of situations where autonomy in machines overshadows the administration of justice in cases of law breaches. There was also consensus that humans must ultimately retain control over advances in AI-made decisions; therefore, the ‘ban’ argument captures the collective dilemma of whether to advocate for outright bans on development or favour stringent regulations that allow for controlled advancement.

The synthesis of the findings of the focus groups and interview studies converge on the AI-driven use of military force as a pivotal catalyst that becomes crucial within discussions on the frontiers of unmanned warfare. The significance of the finding lies in how novel military terms used in the studies, despite implying a need for prior knowledge, did not impede the discourse’s accessibility, marking a notable observation. However, a consensus likely arises from shared values on such a fundamental issue; the undeniable importance of not withdrawing human agency from the use of force stands out.

It is also possible that the participants (incl. experts) may not possess a consistent level of theoretical understanding to notice divergent paradigms regarding the military applications of AI, given that such technology represents a disruptive innovation yet to be fully conceptualised. While it may be the case that an understanding of the intricate technical or strategic details is not uniformly present, this observation intimates a notable deficiency in conceptual thinking regarding emergent AI within multi-domain operations among end-users who might ultimately employ autonomous weapons.

To encapsulate the entire chapter, it is apparent that autonomous weapons come with unique challenges. The idea of an objectively correct and rational output that manifests in a computational course of action must never find a solid and singular basis in military decision-making. Without a doubt, implementing moral values in computational reasoning will be the most severe challenge. However, up to now, it would be misconceived to think of military AI merely as a ‘killer robot’. In theory, autonomous systems in the military domain must be robust and transparent for personnel working with them, along with the awareness of how brittle it is to manipulate the current AI state-of-the-art. Both studies stressed in particular that stakeholders, such as policymakers, the defence industry, and academia, must reach immediate consensus on how the implementation of a regulative framework could be applied and enforced internationally.

Finally, AI weaponisation is garnering increasing attention, and it is becoming evident that comprehensive, in-depth research is essential to keeping pace with technological advancements. The emerging nature of AI in military applications poses unique challenges and questions, many of which are still unexplored. Recognising the limitations of my current study, particularly the constraints posed by the limited sample size, underscores the need for more diverse methodological efforts.

Therefore, I advocate for continuing research, emphasising the necessity to assess the influence of algorithms on the decision-making processes of military leaders, alongside efforts previously undertaken by medical professionals. This endeavour should aim to quantify the extent to which military leaders are prepared to modify their initial judgements based on insights provided by AI. And I leave you with this thought: How can we ensure that soldiers (in the case of Estonia, also reservists) are adequately trained and motivated to operate alongside and counter autonomous systems, fostering a shift in mindset toward the receptivity of human-machine teaming in military operations?

Indeed, such studies are crucial for comprehending the dynamics of human-military AI interaction and guaranteeing the rational integration of artificial autonomy into weapons. However, the decision to adopt AWS remains in human hands. The escalating urgency of the AI arms race should prompt caution in our desires. Echoing the insight of a pioneer (Weizenbaum, 1976), we do not now have ways of making computers wise; we ought not now to give computers tasks that demand wisdom.

REFERENCES

- Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine Ethics* (Illustrated edition). Cambridge University Press.
- AP I, Art 36. (1977). *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*. OHCHR. <https://www.ohchr.org/en/instruments-mechanisms/instruments/protocol-additional-geneva-conventions-12-august-1949-and>
- Asaro, P. (2012). On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886), 687–709. <https://doi.org/10.1017/S1816383112000768>
- Balch, T., & Arkin, R. C. (1998). Behavior-based formation control for multirobot teams. *IEEE Transactions on Robotics and Automation*, 14(6). <https://ieeexplore.ieee.org/document/736776>
- Barrett, L. F. (2021). *This is how your brain makes your mind*. MIT Technology Review. <https://www.technologyreview.com/2021/08/25/1031432/what-is-mind-brain-body-connection/>
- Bednarek, M., & Martin, J. R. (2010). *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*. A&C Black.
- Beecher-Monas, E., & Garcia-Rill, E. (2017). Actus Reus, Mens Rea, and Brain Science: What Do Volition and Intent Really Mean? *Kentucky Law Journal*, 106(2).
- Bergengruen, V. (2024, February 8). How Tech Giants Turned Ukraine Into an AI War Lab. *TIME*. <https://time.com/6691662/ai-ukraine-war-palantir/>
- Bhuta, N., Beck, S., Geiss, R., Liu, H.-Y., & Kress, C. (2016). *Autonomous Weapons Systems: Law, Ethics, Policy*. Cambridge University Press.
- Bo, M., Bruun, L., & Boulanin, V. (2022). *Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS*. Stockholm International Peace Research Institute. <https://doi.org/10.55163/AHBC1664>
- Bode, I., Huelss, H., Nadibaidze, A., Qiao-Franco, G., & Watts, T. F. A. (2023). Prospects for the global governance of autonomous weapons: Comparing Chinese, Russian, and US practices. *Ethics and Information Technology*, 25(1), 5. <https://doi.org/10.1007/s10676-023-09678-x>
- Boothby, W. H., & Heinegg, W. H. von. (2019). *The Law of War: A Detailed Assessment of the US Department of Defense Law of War Manual* (Reprint edition). Cambridge University Press.
- Boylan. (2015). *Public Opinion: A Center of Gravity Leaders Forget* | University of Phoenix Research Hub. <https://research.phoenix.edu/sboylan/publication/public-opinion-center-gravity-leaders-forget>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Cai, H., & Lin, Y. (2012). Coordinating Cognitive Assistance With Cognitive Engagement Control Approaches in Human–Machine Collaboration. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 42(2), 286–294. <https://doi.org/10.1109/TSMCA.2011.2169953>

- Campelo, A. M., & Katz, L. (2020). Older Adults' Perceptions of the Usefulness of Technologies for Engaging in Physical Activity: Using Focus Groups to Explore Physical Literacy. *International Journal of Environmental Research and Public Health*, 17(4), Article 4. <https://doi.org/10.3390/ijerph17041144>
- Cassese, A., Gaeta, P., & Jones, J. (n.d.). The Rome Statute of the International Criminal Court: A Commentary. In *The Rome Statute of the International Criminal Court*. Oxford University Press. Retrieved November 29, 2023, from <https://opil.ouplaw.com/display/10.1093/law/9780198298625.001.0001/law-9780198298625>
- Chadderdon, G. L. (2008). Assessing Machine Volition: An Ordinal Scale for Rating Artificial and Natural Systems. *Adaptive Behavior*, 16(4), 246–263. <https://doi.org/10.1177/1059712308090535>
- Chen, J. (2019). *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2019 International Conference on Human Factors in Robots and Unmanned Systems, July 24–28, 2019, Washington D.C., USA*. Springer.
- Christie, E. H., Ertan, A., Adomaitis, L., & Klaus, M. (2023). Regulating lethal autonomous weapon systems: Exploring the challenges of explainability and traceability. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00261-0>
- Cibralic, B., & Mattingly, J. (2022). Machine agency and representation. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01446-7>
- Clarkson, C. M. V. (2005). *Understanding Criminal Law*. Sweet & Maxwell.
- Clausewitz, C. von. (2004). *On War*. Taylor & Francis.
- Coeckelbergh, M. (2020). *AI Ethics*. The MIT Press.
- Colorio, M. (2021). A Commander's Motivations and Geographical Remoteness under Command Responsibility: An Analysis of Controversial Issues of the Bemba Appeal Judgment. *International Criminal Law Review*, 21(3), 445–475. <https://doi.org/10.1163/15718123-bja10072>
- Cotterrell, R. (2018). *Sociological Jurisprudence: Juristic Thought and Social Inquiry*. <https://www.routledge.com/Sociological-Jurisprudence-Juristic-Thought-and-Social-Inquiry/Cotterrell/p/book/9781138052840>
- Dai, T., & Singh, S. (2023). *Artificial Intelligence on Call: The Physician's Decision of Whether to Use AI in Clinical Practice* (SSRN Scholarly Paper 3987454). <https://doi.org/10.2139/ssrn.3987454>
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What Is a Knowledge Representation? *AI Magazine*, 14(1), Article 1. <https://doi.org/10.1609/aimag.v14i1.1029>
- Defence AI Strategy. (2022). *Ambitious, safe, responsible: Our approach to the delivery of AI-enabled capability in Defence*. Defence Artificial Intelligence Centre. <https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence>
- Del Monte, L. A. (2018). *Genius Weapons: Artificial Intelligence, Autonomous Weaponry, and the Future of Warfare* (Illustrated edition). Prometheus.
- Directive 3000.09, Autonomy in Weapon Systems. (2012). *DoD Directive 3000.09, Autonomy In Weapon Systems*. U.S. Department of Defense. <https://www.defense.gov/News/Releases/Release/Article/3278076/dod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems/https%3A%2F%2Fwww.defense.gov%2FNews%2FReleases%2FRelease%2FArticle%2F3278076%2Fdod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems%2F>

- Donnelly-Lazarov, B. (Ed.). (2015). Actus reus and mens rea. In *A Philosophy of Criminal Attempts* (pp. 107–133). Cambridge University Press. <https://doi.org/10.1017/CBO9781139342636.007>
- Dresp-Langley, B. (2023). The weaponization of artificial intelligence: What the public needs to be aware of. *Frontiers in Artificial Intelligence*, 6. <https://www.frontiersin.org/articles/10.3389/frai.2023.1154184>
- Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press.
- Evans, K. D., Robbins, S. A., & Bryson, J. J. (2023). Do We Collaborate With What We Design? *Topics in Cognitive Science*, n/a(n/a). <https://doi.org/10.1111/tops.12682>
- Farr, R. (1993). *Common sense, science and social representations*. <https://doi.org/10.1088/0963-6625/2/3/001>
- Fleck, D. (2021). The Handbook of International Humanitarian Law. In *The Handbook of International Humanitarian Law*. Oxford University Press. <https://opil.ouplaw.com/display/10.1093/law/9780198847960.001.0001/law-9780198847960>
- Galliot, J., MacIntosh, D., & Ohlin, J. D. (2021). *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*. New York: Oxford University Press.
- GGE on LAWS. (2023). *Convention on Certain Conventional Weapons -Group of Governmental Experts on Lethal Autonomous Weapons Systems (2023) | United Nations*. <https://meetings.unoda.org/ccw-/convention-on-certain-conventional-weapons-group-of-governmental-experts-on-lethal-autonomous-weapons-systems-2023>
- Goldstein, B. (2008). Lost in Translation: Some Brief Notes on Writing about Law for the Layperson. *New York Law School*.
- Gómez de Ágreda, A. (2020). Ethics of autonomous weapons systems and its applicability to any AI systems. *Telecommunications Policy*, 44(6), 101953. <https://doi.org/10.1016/j.telpol.2020.101953>
- Goussac, N., Boulanin, V., & Carlsson, M. P. (2020). *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*. SIPRI. <https://www.sipri.org/publications/2020/other-publications/limits-autonomy-weapon-systems-identifying-practical-elements-human-control-0>
- Granillo, A. (2021). *NATOC2COE – Nato Command and Control Centre of Excellence*. <https://c2coe.org/wp-content/uploads/Library%20Documents/Study/20210430%20MDC%20Long%201.0.pdf>
- Gray, M., & Ertan, A. (2021). *Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment*. CCDCOE, NATO. https://ccdcoe.org/uploads/2021/12/Strategies_and_Deployment_A4.pdf
- Heller, K. J. (2023). *The Concept of “The Human” in the Critique of Autonomous Weapons* (SSRN Scholarly Paper 4342529). <https://doi.org/10.2139/ssrn.4342529>
- Hensel, H. M. (2016). *The Legitimate Use of Military Force: The Just War Tradition and the Customary Law of Armed Conflict*. Routledge.
- Herring, J. (2018). *Criminal Law: Text, Cases, and Materials*. Oxford University Press.
- Hirsh, M. (2024, February 22). How AI Will Revolutionize Warfare. *Foreign Policy*. <https://foreignpolicy.com/2023/04/11/ai-arms-race-artificial-intelligence-chatgpt-military-technology/>
- Horowitz, M. C. (2019). When speed kills: Lethal autonomous weapon systems, deterrence and stability. *Journal of Strategic Studies*, 42(6), 764–788. <https://doi.org/10.1080/01402390.2019.1621174>

- Horowitz, M. C., Kahn, L., Macdonald, J., & Schneider, J. (2023). Adopting AI: How familiarity breeds both trust and contempt. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-023-01666-5>
- Hunt, A. (1986). Review of A Sociological Theory of Law [Review of *Review of A Sociological Theory of Law*, by N. Luhmann, E. King-Utz, & M. Albrow]. *The British Journal of Criminology*, 26(2), 192–194. <https://www.jstor.org/stable/23636759>
- iMUGS. (2020). *EDIDP: The iMUGS Consortium demonstrates autonomous missions with robotic systems*. https://defence-industry-space.ec.europa.eu/edidp-imugs-consortium-demonstrates-autonomous-missions-robotic-systems-2022-12-19_en
- Jørgensen, M. W., & Phillips, L. J. (2002). *Discourse Analysis as Theory and Method*. SAGE Publications.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7–15. [https://doi.org/10.1016/0020-0190\(89\)90102-6](https://doi.org/10.1016/0020-0190(89)90102-6)
- Kanellakis, C., & Nikolakopoulos, G. (2017). Survey on Computer Vision for UAVs: Current Developments and Trends. *Journal of Intelligent & Robotic Systems*, 87(1), 141–168. <https://doi.org/10.1007/s10846-017-0483-z>
- Keller, M., & Kalmus, V. (2009). What makes me cool? Estonian tweens' interpretative repertoires. *Young Consumers: Insight and Ideas for Responsible Marketers*, 10, 329–341. <https://doi.org/10.1108/17473610911007157>
- Kleykamp, M., & Yastrzemsky, J. R. (2017). Military Sociology. In K. O. Korgen (Ed.), *The Cambridge Handbook of Sociology: Core Areas in Sociology and the Development of the Discipline* (Vol. 1, pp. 344–354). Cambridge University Press. <https://doi.org/10.1017/9781316418376.035>
- Koch, W. (2022). Elements of an Ethical AI Demonstrator for Responsibly Designing Defence Systems. *2022 25th International Conference on Information Fusion (FUSION)*, 1–8. <https://doi.org/10.23919/FUSION49751.2022.9841387>
- Korteling, J. E. (Hans), van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4. <https://www.frontiersin.org/articles/10.3389/frai.2021.622364>
- Krippendorff, K., & Bock, M. A. (2008). *The Content Analysis Reader* (1st edition). SAGE Publications, Inc.
- Krueger, A. E., Pollmann, K., Fronemann, N., & Foucault, B. (2020). Guided User Research Methods for Experience Design – A New Approach to Focus Groups and Cultural Probes. *Multimodal Technologies and Interaction*, 4(3), Article 3. <https://doi.org/10.3390/mti4030043>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), Article 7553. <https://doi.org/10.1038/nature14539>
- Lee, C.-E., Baek, J., Son, J., & Ha, Y.-G. (2023). Deep AI military staff: Cooperative battlefield situation awareness for commander's decision making. *The Journal of Supercomputing*, 79(6), 6040–6069. <https://doi.org/10.1007/s11227-022-04882-w>
- Lenharo, M. (2023). An AI revolution is brewing in medicine. What will it look like? *Nature*, 622(7984), 686–688. <https://doi.org/10.1038/d41586-023-03302-0>
- Leveringhaus, A. (2021). *Autonomous Weapons and the Future of Armed Conflict*. In *Lethal Autonomous Weapons*. Oxford University Press. <https://doi.org/10.1093/oso/9780197546048.003.0012>
- Liao, S. M. (Ed.). (2020). *Ethics of Artificial Intelligence* (1st edition). Oxford University Press.

- Lin, P., Abney, K., & Jenkins, R. (Eds.). (2017). *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press. <https://doi.org/10.1093/oso/9780190652951.001.0001>
- Lucas, G., Syse, H., Helkala, K., & Barrett, E. (2023). *Ethical Challenges in AI-enhanced Military Operations*. Frontiers Media SA.
- Lundberg, J., & Johansson, B. J. E. (2021). A framework for describing interaction between human operators and autonomous, automated, and manual control systems. *Cognition, Technology & Work*, 23(3), 381–401. <https://doi.org/10.1007/s10111-020-00637-w>
- Macrina, G., Di Puglia Pugliese, L., Guerriero, F., & Laporte, G. (2020). Drone-aided routing: A literature review. *Transportation Research Part C: Emerging Technologies*, 120, 102762. <https://doi.org/10.1016/j.trc.2020.102762>
- Maslej et al. (2023). *The AI Index 2023 Annual Report*. The AI Index 2023 Annual Report. <https://aiindex.stanford.edu/report/>
- Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), Article 2. <https://doi.org/10.17169/fqs-1.2.1089>
- McFarland, T. (Ed.). (2020). Introduction. In *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law* (pp. 1–7). Cambridge University Press. <https://doi.org/10.1017/9781108584654.001>
- McFarland, T., & Assaad, Z. (2023). Legal reviews of in situ learning in autonomous weapons. *Ethics and Information Technology*, 25(1), 9. <https://doi.org/10.1007/s10676-023-09688-9>
- Meerveld, H. W., Lindelauf, R. H. A., Postma, E. O., & Postma, M. (2023). The irresponsibility of not using AI in the military. *Ethics and Information Technology*, 25(1), 14. <https://doi.org/10.1007/s10676-023-09683-0>
- Meyer, J., Khademi, A., Têtu, B., Han, W., Nippak, P., & Remisch, D. (2022). Impact of artificial intelligence on pathologists' decisions: An experiment. *Journal of the American Medical Informatics Association: JAMIA*, 29(10), 1688–1695. <https://doi.org/10.1093/jamia/ocac103>
- Mitchell, M. (2020). *Artificial Intelligence: A Guide for Thinking Humans*. Pelican.
- Moelker, R., & Schenk, N. (2018). Mixing up Humans and Military Technology. In G. Caforio & M. Nuciari (Eds.), *Handbook of the Sociology of the Military* (pp. 443–462). Springer International Publishing. https://doi.org/10.1007/978-3-319-71602-2_23
- Morgan, D. (1996). Focus Groups. *Annual Review of Sociology*, 22, 129–152. <https://doi.org/10.1146/annurev.soc.22.1.129>
- Morgan, F. E., Boudreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K., & Grossman, D. (2020). *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR3139-1.html
- Moscovici, S. (2001). *Social Representations: Explorations in Social Psychology*. NYU Press.
- Mozaffari, M., Saad, W., Bennis, M., Nam, Y.-H., & Debbah, M. (2019). A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Communications Surveys & Tutorials*, 21(3), 2334–2360. <https://doi.org/10.1109/COMST.2019.2902862>
- Nyholm, S. (2018). Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics*, 24(4), 1201–1219. <https://doi.org/10.1007/s11948-017-9943-x>

- Parent, M., Gallupe, R. B., Salisbury, Wm. D., & Handelman, J. M. (2000). Knowledge creation in focus groups: Can group technologies help? *Information & Management*, 38(1), 47–58. [https://doi.org/10.1016/S0378-7206\(00\)00053-7](https://doi.org/10.1016/S0378-7206(00)00053-7)
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide* (pp. xv, 336). Blackwell Publishing. <https://doi.org/10.1002/9780470754887>
- Posit team. (2023). *Posit*. Posit. <https://www.posit.co/>
- Purves, D., Jenkins, R., & Strawser, B. J. (2015). Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 18(4), 851–872. <https://doi.org/10.1007/s10677-015-9563-y>
- Ratinaud, P. (2023). *Iramuteq—IRaMuTeQ*. <http://www.iramuteq.org/>
- Roff, H. M. (2013). Killing in War: Responsibility, liability, and lethal autonomous robots. In *Routledge Handbook of Ethics and War*. Routledge.
- Rosert, E., & Sauer, F. (2019). Prohibiting Autonomous Weapons: Put Human Dignity First. *Global Policy*, 10(3), 370–375. <https://doi.org/10.1111/1758-5899.12691>
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Books.
- Russell, S. (2023). AI weapons: Russia’s war in Ukraine shows why the world must enact a ban. *Nature*, 614(7949), 620–623. <https://doi.org/10.1038/d41586-023-00511-5>
- Rutchick, A. M., McManus, R. M., Barth, D. M., Youmans, R. J., Ainsworth, A. T., & Goukassian, H. J. (2017). Technologically facilitated remoteness increases killing behavior. *Journal of Experimental Social Psychology*, 73, 147–150. <https://doi.org/10.1016/j.jesp.2017.07.001>
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5. <https://www.frontiersin.org/articles/10.3389/frobt.2018.00015>
- Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Saxon, D. (2021). *Fighting Machines: Autonomous Weapons and Human Dignity*. University of Pennsylvania Press.
- Scanlan, C. L. (2020). *Preparing for the Unanticipated: Challenges in Conducting Semi-Structured, In-Depth Interviews*. SAGE Publications Ltd. <https://doi.org/10.4135/9781529719208>
- Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. W. W. Norton & Company.
- Scharre, P. (2023). *Four Battlegrounds: Power in the Age of Artificial Intelligence*. W. W. Norton & Company.
- Scheufele, D. A., Brossard, D., & Newman, T. (2023, September 5). *Experts alone can’t handle AI – social scientists explain why the public needs a seat at the table*. The Conversation. <http://theconversation.com/experts-alone-cant-handle-ai-social-scientists-explain-why-the-public-needs-a-seat-at-the-table-210848>
- Schmitt, M. N., & Goddard, D. S. (2016). International law and the military use of unmanned maritime systems. *International Review of the Red Cross*, 98(902), 567–592. <https://doi.org/10.1017/S1816383117000339>
- Schmitt, M. N., Waxman, M. C., & Oakley, T. W. (2022). *The Future Law of Armed Conflict*. Oxford University Press.
- Schwarz, E. (2021). Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control. *Philosophical Journal of Conflict and Violence*, 5(1), 53–72. <https://doi.org/10.22618/TP.PJCV.20215.1.139004>

- SFA23. (2023). *Strategic Foresight Analysis 2023| NATO ACT*. NATO's ACT. https://www.act.nato.int/wp-content/uploads/2024/01/SFA2023_Final.pdf
- Sharkey, A. (2019). Autonomous weapons systems, killer robots and human dignity. *Ethics and Information Technology*, 21(2), 75–87. <https://doi.org/10.1007/s10676-018-9494-0>
- Shields, P. M. (2020). Dynamic Intersection of Military and Society. In A. Sookermany (Ed.), *Handbook of Military Sciences* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-030-02866-4_31-1
- Simpson, T. W., & Müller, V. C. (2016). Just War and Robots' Killings. *The Philosophical Quarterly*, 66(263), 302–322. <https://doi.org/10.1093/pq/pqv075>
- Singer, P. (Ed.). (1993). *A Companion to Ethics* (1st edition). Wiley-Blackwell.
- SIPRI. (2021). *SIPRI Yearbook 2021: Armaments, Disarmament and International Security*. Oxford University Press.
- Smith, P. T. (2022). Resolving responsibility gaps for lethal autonomous weapon systems. *Frontiers in Big Data*, 5. <https://www.frontiersin.org/articles/10.3389/fdata.2022.1038507>
- Soltanzadeh, S. (2022). Strictly Human: Limitations of Autonomous Systems. *Minds and Machines*, 32(2), 269–288. <https://doi.org/10.1007/s11023-021-09582-7>
- Sourati, J., & Evans, J. A. (2023). Accelerating science with human-aware artificial intelligence. *Nature Human Behaviour*, 7(10), Article 10. <https://doi.org/10.1038/s41562-023-01648-z>
- Sparrow, R. (2016). Robots and Respect: Assessing the Case Against Autonomous Weapon Systems. *Ethics and International Affairs*, 30(1), 93–116. <https://doi.org/10.1017/s0892679415000647>
- Strawser, B. J. (2013). *Killing by Remote Control: The Ethics of an Unmanned Military*. OUP USA.
- Sutrop, M. (2020). DEEP CONCEPTUAL MORAL DISAGREEMENTS: OVER WHAT DO WE DISAGREE AND WHY? *Trames. Journal of the Humanities and Social Sciences*, 24(3), 295. <https://doi.org/10.3176/tr.2020.3.03>
- Sztompka, P. (1999). *Trust: A Sociological Theory*. Cambridge University Press.
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Penguin UK.
- Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). *Methods of Text and Discourse Analysis: In Search of Meaning* (First edition). SAGE Publications Ltd.
- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, 4(6), 479–491. <https://doi.org/10.1017/S1930297500004022>
- van de Poel, I., Frank, L. E., Hermann, J., Hopster, J., Lenzi, D., Nyholm, S., Taebi, B., & Ziliotti, E. (2023). *Ethics of Socially Disruptive Technologies: An Introduction*. Open Book Publishers. <https://doi.org/10.11647/obp.0366>
- van Dyck, L., Kwitt, R., Denzler, S., & Gruber, W. (2021). Comparing Object Recognition in Humans and Deep Convolutional Neural Networks – An Eye Tracking Study. *Frontiers in Neuroscience*, 15, 750639. <https://doi.org/10.3389/fnins.2021.750639>
- Vaportzis, E., Giatsi Clausen, M., & Gow, A. J. (2017). Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. *Frontiers in Psychology*, 8, 1687. <https://doi.org/10.3389/fpsyg.2017.01687>

- Verdiesen, I., Santoni de Sio, F., & Dignum, V. (2021). Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight. *Minds and Machines*, 31(1), 137–163. <https://doi.org/10.1007/s11023-020-09532-9>
- Wagner, W., & Talves, K. (2023). Integrated modular unmanned ground systems: Social, Professional, Ethical, and Legal Aspects. *Academia Militaris*.
- Wagner, W., Viidalepp, A., Idoiaga-Mondragon, N., Talves, K., Lillemäe, E., Pekarev, J., & Otsus, M. (2023). Lay representations of artificial intelligence and autonomous military machines. *Public Understanding of Science*, 09636625231167071. <https://doi.org/10.1177/09636625231167071>
- Wareham, M. (2020). Stopping Killer Robots. *Human Rights Watch*. <https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and>
- Weizenbaum, J. (1966). *Contextual Understanding by Computers*. (Vol. 9).
- Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman.
- Wetherell, M., & Potter, J. (1992). *Mapping the Language of Racism: Discourse and the Legitimation of Exploitation*. Columbia University Press.
- Xie, H. (2023). The promising future of cognitive science and artificial intelligence. *Nature Reviews Psychology*, 2(4), Article 4. <https://doi.org/10.1038/s44159-023-00170-3>
- Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-22244-7>
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, 9(4), Article 4. <https://doi.org/10.1007/s13244-018-0639-9>
- Yeşilbaş, İ., & Çetin, Ş. (2019). Trust in military leader: Scale development and validation. *Military Psychology*, 31(2), 147–159. <https://doi.org/10.1080/08995605.2019.1578150>
- Young, S. H., Mazzuchi, T. A., & Sarkani, S. (2017). A Framework for Predicting Future System Performance in Autonomous Unmanned Ground Vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7), 1192–1206. <https://doi.org/10.1109/TSMC.2016.2563403>
- Zajac, M. (2023). AWS compliance with the ethical principle of proportionality: Three possible solutions. *Ethics and Information Technology*, 25(1), 13. <https://doi.org/10.1007/s10676-023-09689-8>
- Zeng, Y., Zhang, R., & Lim, T. J. (2016). Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Communications Magazine*, 54(5), 36–42. <https://doi.org/10.1109/MCOM.2016.7470933>

SUMMARY IN ESTONIAN

Sõjaline tehisintellekt ja autonoomsed relvasüsteemid: interdistsiplinaarne uurimus eetilistest, õiguslikest ja sotsioloogilistest aspektidest

Autonoomseid relvasüsteeme, mida on nimetatud ka tapjarobotiteks, võib õigustatult pidada järgmiseks revolutsiooniliseks verstapostiks sõjateaduses pärast püssirohu ja tuumarelva leiutamist. Tehisintellekti potentsiaali rakendamine sõjalistel eesmärkidel, nii otsustusprotsessis kui mehitamata platvormide kasutamisel, tähistab kahtlemata murrangulist hetke sõjalise jõu kasutamises, mis on traditsiooniliselt olnud inimesekeskne. Autonoomsete relvasüsteemide arendamine annab alust arvata, et sõjaline tehisintellekt seab kaalule inimelud. Elu ja surma puudutavate otsuste delegeerimine masinatele on käivitanud eksistentiaalse sisuga debati, mille keskmes on eetilised, õiguslikud ja sotsioloogilised kategooriad.

Teaduskirjanduses keskendutakse autonoomsete relvasüsteemide rakendamisel peamiselt kahele tõekspidamisele. Ühelt poolt vähendavad tänapäevased relvajõud inimeste kohalolekut lahinguväljal ning toetuvad üha enam mehitamata platvormidele, et täita monotoonseid või ohtlikke ülesandeid pikema aja vältel. Nii väidetakse, et mehitamata sõjapidamise eeliseks on inimlike füsioloogiliste ja psühholoogiliste piirangute puudumine. Teisisõnu, masinad on väsimatud, ei tunne hirmu, ei alistu ning ei vaja motiveerimist ega juhtimisstruktuuri. Teiselt poolt ollakse väga selgelt algoritmilise sõjapidamise vastu, sest tehisintellekti relvastamisel tuuakse välja potentsiaalsed ohud, mis tulenevad sellest, et masinate autonoomne käitumine võib olla ettearvamatu. Ennekõike arvatakse, et sõdurite eemaldamine otsesest lahingutegevusest võib muuta regulatsioonidest kinnipidamise masinlikult sirgjooneliseks. See tähendab, et autonoomsete relvade langetatud otsused võivad olla sõjaliselt ratsionaalsed, kuid mitte tingimata humaansed, õiguslikult kaalutletud ega ühiskondlikult vastuvõetavad.

Delegeerides sõjalise jõu inimeselt masinale keskendub vaidlus sellele, kui tõhusalt suudavad autonoomsed relvasüsteemid järgida rahvusvahelise humanitaarõiguse põhiprintsiipe. Vaidlusalused printsiibid hõlmavad legaalse sihtmärgi eristamist, sõjalise vajaduse hindamist vastase lahinguvõime vähendamiseks ning proportsionaalsuse säilitamist kaasneva kahju ja saavutatava sõjalise eelise vahel. Siinjuures tuleb tõdeda, et nende printsiipide rakendamine kaas-aegses relvakonfliktis on väljakutseks isegi kogenud sõjaväelastele ja relvajõudude õigusnõunikele.

Relvakonflikti on iseloomustatud ka kui sõjaudu, kus segunevad võimalused, tõenäosus, õnn ja inimlik tahtevabadus. Tehisintellekti treenimine sellises kaoses tegutsemiseks on tõsine väljakutse, mille tulemuslikkust on keeruline ette näha. Siiski peab ütleva, et ka sõdurite käitumine võib olla ettearvamatu, kuna sõjakuritegusid panevad toime ikkagi inimesed, mitte masinad. Paraku iseloomustab samasugune ettearvamatus ka masinlikku autonoomiat, kuna tehnoloogiat võivad

haavata kallutatud algoritmid ja küberrünnakud, põhjustades soovimatuid tagajärgi.

Doktoritöös uurisin sõjateaduse ja -tehnoloogia (relvasüsteemid) ning arvuti-teaduse (tehisintellekti autonoomia) ristumiskohti. Nende teadusalade paradigmatel koosmõju uurimiseks oli keeruline leida sobivat lähenemisviisi. Kaalule jäi interdistsiplinaarne vaatenurk, mis annab ülevaate enim käsitlemist leidnud eetilistest, õiguslikest ja sotsioloogilistest aspektidest tehisintellekti relvastamise debatis. Doktoritöö eesmärk oli teada saada, mille üle debateeritakse algoritmilises sõjapidamises kõige enam, ja saada uusi teadmisi, analüüsides tavakodanike ja militaareksperptide hinnanguid sõjalise jõu ja tehisintellekti integreerimisele. Nimetatud eesmärki silmas pidades koostas süsteemse teaduskirjanduse ülevaate (**Uuring I**), et mõista debati argumentatsiooni ja välja töötada intervjuuküsimustik nende teemade analüüsimiseks.

Osa uuringuks vajalikke andmeid kogusin mehitamata maismaasüsteeme uurivas teadusrühmas (projekt iMUGS). Avaliku arvamuse kuulamata jätmine niivõrd tundlikul teemal risk ulatusliku vastuseisu tekkimiseks ühiskonnas, kuna inimeste mõtted ja hoiakud tehnoloogiliste uuenduste osas on nende rakendamisel olulised. Fookusgruppide uuring (**Uuring II**) keskendus inimeste mõtete ja hoiakute analüüsimisele seoses mehitamata sõjapidamisega. Samavõrd oluline oli ära kuulata ka sõjalised eksperdid: tegin intervjuud tegevväelaste ja lõppkasutajatest teenistujatega (**Uuring III**), et analüüsida ekspertide hinnanguid tehisintellekti rolli kohta kaitsevæes.

Pärast põhjalikku tutvumist temaatilise teaduskirjandusega ja kahe uuringu läbiviimist, keskendusin sõjalise tehisintellekti ja autonoomsete süsteemide integreerimisega seotud keerukate väljakutsete uurimisele. Resümeeeritud seisukohad ja soovitused tulevasteks uuringuteks esitan siinse kokkuvõtte lõpus. Esmalt tutvustan teadustöö tulemusi, vastates sissejuhatuses esitatud uurimisküsimustele.

Uurimisküsimus 1: Millised on autonoomsete relvasüsteemide kasutuselevõtuga seotud peamised aspektid teaduskirjanduses? (**Uuring I**)

Seoses autonoomsete relvasüsteemide kasutuselevõtuga käsitletakse kirjanduses kõige enam nende võimekust kohaldada autonoomselt humanitaarõiguse printsiipe. Sellele järgneb vastutajate küsimussüütegude omistamise kohta seoses autonoomiast tulenevate võimalike tagajärgedega. Lisaks kerkivad teemadena esile elu ja surma puudutavate otsuste delegeerimise eetilisus, inimesele arusaadava kontrolli tagamine tehisintellekti toimepidevusest, dehumaniseerimine sihtmärgistamisel, inimväärikusele väiksema tähtsuse omistamine, võidurelvastamise vältimine ja relvastatud konflikti lävendi vähenemine.

a. Mis on nende aspektide aluseks olev argumentatsioon? (**Uuring I**)

Debati kese seisneb argumentatsioonis, kas tehisintellekti juhitud süsteemid on võimelised paremini või halvemini järgima humanitaarõiguse põhimõtteid kui professionaalsed sõjaväelased. Ühest küljest ei ole arvutused eeltreenitud andmete põhjal piisavad või on piiratud, et tõlgendada printsiipide kohaldamist

elulistele asjaoludele. Teisest küljest on diskussioon vastupidine, kuna masin töötleb infot erapooletult ilma emotsioonideta, eelarvamusteta, inimlike füüsiliste nõrkusteta, mis võivad andmeid kallutada või sobitada infot väljakujunenud veendumuste või väärtustega. Märkimisväärne osa argumentatsioonist oli seotud õiguspärasusega, nt kas sõjalise jõu delegeerimine masinale on legitiimne. Siinjuured kerkib esile vastutuse küsimus, mis viitab tehnikule autonoomiale kui süüd välistavale asjaolule. Muuhulgas näitas süstemaatiline analüüs, et ligikaudu pool väitlusest tugines normatiivsele eetikale, mille seas prevaleeris deontoloogiline positsioon. Näiteks seisukoht, et elu ja surma puudutavad otsused väljaspool inimkontrolli on juba olemuslikult valed, peaks olema piisav põhjus, et tehiskontrolli relvastamine peatada.

Uurimisküsimus 2: Mis moodustab autonoomse relvasüsteemi ja kuidas mõista relva autonoomiat? (**Uuring II–III**).

Nii nagu autonoomne relvasüsteem ise, on ka selle kohta käivad erinevad definitsioonid kujunemisjärgus. Laiemas kontekstis võib tehiskontrolli rakendamist vaadelda kui sõjalise jõu mitmekordistajat, mis tajub masinlike sensoritega ümbritsevat keskkonda, seostab mingil määral olukorra muutuseid, tuvastab sihtmärke ja on seejärel võimeline tegema otsuseid surmava jõu kasutamiseks ilma inimese sekkumiseta. Käesolevas teadustöös seostub autonoomia ennekõike inimeseks olemise vaba tahtega teha iseseisvaid otsuseid, mida on keeruline samaväärselt omistada masinale. Nii on relvasüsteemi autonoomia hüpoteetiline omadus. Teadaolevalt ei eksisteeri algoritmi kujul, mis võimaldaks relval toimida olukorradeadlikult ilma väliste kontrollimehhanismideta. Dissertatsiooni aluseks on eeldus, et sõjalise tehiskontrolli autonoomia võib areneda viisil, mis võimaldab süsteemil kohaneda olukorraga, iseseisvalt tuvastada sõjalisi sihtmärke ja langetada otsuseid surmava jõu kasutamise kohta, arvestades sõjalist vajadust ja proportsionaalsust. Siinjuures on oluline mõista, kui pidada eelpool kirjeldatud stsenaariumi ebatõenäoliseks või selline areng üldse välistada, siis ei oma ka sõjalise tehiskontrolli eetiliste, õiguslike ja sotsioloogiliste aspektide analüüsimine tähtsust.

a. Millised on uuringus osalenute hinnangud masinate ja inimeste tehtud otsuste kohta seoses elu ja surma puudutavate küsimustega? (**Uuring II–III**)

Otsustusprotsessi iseloomustavad erinevad kaalutlused. Osalejad kaldusid uskuma, et eetilisel tundlike otsuste tegemine on omane ainult inimestele, kuna masinatele puudub moraalne kompass. Samuti toodi esile, et masinatele puudub olukorradeadlikkust, mis võimaldab korrigeerida sündmuste kulgu. Vastustes rõhutatakse (**Uuring II**), et hinnangu andmine olukorrale sõltub seetõttu otsuste tegemisele rakendatavast perspektiivist. Lisaks selgub tulemustest (**Uuring III**), et ekspertide arvates on otsustusprotsess olemuslikult sõltuv kontekstist. Informatsiooni on tihtipeale vähe ja seepärast on otsuse tegemiseks oluline infokilde ja fakte tõlgendada ja seostada ning mõista neid seoseid laiemas kontekstis. Nõnda leiti konsensuslikult (**Uuring III**), et inimese sekkumine masina konstrueeritud

tegevuskäikudesse on kriitilise tähtsusega eetilisel tundlike otsuste tegemisel. Seda ennekõike seepärast, et otsused nõuavad lisaks objektiivsele analüüsile ka moraalset hinnangut, milleks ei ole masinõppel põhinevad süsteemid hetkel võimelised.

- b. Kes peaks uuringus osalenute arvates vastutama, kui autonoomsete relvade kasutamine on põhjustatud sõjakuritegusid? Miks võib sellistel juhtudel esineda lünk õigusmõistmises? (**Uuring II–III**)

Mõlemas uuringus valitses konsensus, et sõdur, masin või muu käsutäitja on vahend, välistades ühtlasi süsteemi programmeerija ja tootja vastutuse. Isik, kes annab käsu autonoomse relva kasutamiseks, peaks vastutama süütegude eest, mis tulenevad sellest korraldusest. Mehitamata platvormide põhjustatud õigusvastaste tagajärgede esinemisel tehti sellegipoolest mõõndus, et süüteo episoodides võib olla keeruline tuvastada teo toimepanija tahtlus tehisintellekti ettearvamatu käitumise tõttu. Küll aga jagasid osalejad seisukohta, et võimalikku lünka õigusmõistmises peab ennetama ja vastutus peab olema eelnevalt reguleeritud.

- c. Miks peavad uuringus osalejad autonoomsete relvade arendamist eetilisel õigeks või valeks? Millised on põhjused arendustegevuse jätkamiseks või keelustamiseks? (**Uuring II–III**)

Autonoomsete relvade arendamine ei ole eetilisel õige ega vale. Uuringud toovad esile, et see sõltub suurel määral kasutamise eesmärgist. Autonoomsete relvade täielik keelustamine oli osalejate arvates ebatõenäoline. Seetõttu leiti, et vajalike meetmete kehtestamine ja standardite väljatöötamine murranguliste tehnoloogiate arendamiseks on mõõdapääsmatu. Kokkuvõtlikult võib öelda, et uuringus osalejate arvates on algoritmiline sõjapidamine tulnud, et jääda, kuid tehnoloogiline innovatsioon ja selle potentsiaalsed eetilised, juriidilised ja ühiskondlikud mõjud peavad olema tasakaalus.

- d. Milliseid dilemmasid väljendavad uuringus osalejad autonoomsete relvasüsteemide kohta? (**Uuring II**)

Osalejad leidsid peamiselt, et autonoomsed süsteemid on riigi kaitsmise eesmärgil eelistatavad kui traditsiooniline kaitsevägi, kuna see säästab kaasmaalaste elusid. Samas eelistasid osalejad ise võidelda vastase sõdurite, mitte robotite vastu, sest inimese käitumine on etteaimatavam ja vastasega kohtumine vähem hirmutav kui vaenulike mehitamata platvormidega. Kirjeldatud eelistused illustreerivad paradoksaalseid valikuid, sest sarnaseid süsteeme võib kasutada mõlemal poolel ja mingil hetkel peavad inimesed lahinguväljal siiski robotitega silmitsi seisma.

Uurimisküsimus 3: Milline on ekspertide hinnangul autonoomsete relvasüsteemide prognoositav mõju sõjaliste operatsioonidele? (**Uuring III**)

Ekspertid leiavad, et tehisintellektil põhinev sõjaline otsustusprotsess on väga problemaatiline, kuna isegi professionaalsetel ja kogenud juhtidel on piiratud aja ja info tingimustes keeruline otsuseid langetada. Valdavalt oldi seisukohal, et mehitamata platvormide kasutamine ei muuda tingimata sõjakunsti olemust. Selle aluseks on tõdemus, et sõda on jäädavalt inimeste vaheline võitlus ja oma tahtejõu peale surumine ning masinatel puudub võitluseks vaba tahe. Osalejad rõhutasid, et patriootlik mentaliteet ja vabaduse eest võitlemise kaitsetahe on midagi, mida ei saa algoritmidesse kodeerida. Ekspertid järeldasid, et mehitamata platvormid ei ole veel võimelised täitma relvajõudude põhifunktsioone ning nende käigushoidmine nõuab sageli märkimisväärset inimressurssi. See tähendab, et vastupidiselt levinud arvamusele, autonoomsed süsteemid ei pruugi vabastada sõjaväelasi oluliselt nende põhiülesannetest, vaid võivad suurendada tegevälaste kaasamise vajadust süsteemide efektiivseks opereerimiseks. Sõjalise tehisintellekti integreerimist reservvälastest koosnevasse sõjaaja koosseisu nähakse ennatlikuna, sest esmalt tuleb tagada kaasaegse individuaalse ja konventsionaalse varustuse ja relvastuse olemasolu. Ekspertid on veendunud, et lahingutegevuse põhiraskust kannavad jätkuvalt sõdurid, mitte relvastatud tehisintellekti süsteemid. Samas ei välista ükski ekspert autonoomsete tehnoloogiate potentsiaalseid eeliseid, vaid rõhutatakse mehitamata platvormide ja välja õpetatud reservvälaste vahelist tasakaalu.

a. Mis põhjustab ebakindlust autonoomsete süsteemide kasutamisel? (**Uuring III**)

Ebakindlus tuleneb peamiselt teadmatusest, kas algoritmid suudavad jäljendada sõjalise otsustusprotsessi detailsust. Sõjalise sihtmärgi eristamine on niivõrd nüansirikas, näiteks hoonestatud aladel, kus vastane on segunenud tsiviilelanikkonnaga. Kuigi algoritmidel puuduvad inimlikud nõrkused ja vajadused, mis teevad rasketes oludes jätkusuutlikeks, valitses ekspertide seas konsensus, et inimkontrolli säilitamine elu ja surma diskretsioonis on asendamatu. Lisaks võib tehisintellektiga konsulteerimine kaasa tuua moraalseid dilemmasid, näiteks süsteem tuvastab sõjalise sihtmärgi, kuid mehitatud üksus samas asukohas seda sihtmärki ei kinnita või vastupidi. Otsustajad satuvad situatsiooni, milles on keeruline hinnata, milliseid andmeid usaldada. Selline olukord ei ole ainult üksik kaasus, vaid näitab vajadust täiustatud protseduuride ja lahingutegevuse reeglite järele, mis aitaksid määratleda, kuidas toimida tehisintellekti ja inimese poolt vahendatud informatsiooni ja hinnangute vastukäivuse korral.

b. Mis on tundmatu suurus mehitamata sõjapidamises ja kuidas seda leevendada? (**Uuring III**)

See, kuidas autonoomsed süsteemid jõuavad tegevuskäikudeni, ei ole lõppkasutajatele läbipaistev, mistõttu leidsid intervjueeritavad, et täiendava tehnilise personali ja õigusnõunike kaasamine tundub hädavajalik. Samas muudab isikkoosseisu suurendamine juhtimise aeganõudvaks ja juhtimispunkti kergemini

avastatavaks. Mehitatud üksuste ning mehitamata platvormide tegevuse sünkroniseerimine muutub samuti keerulisemaks, mis omakorda pärsib situatsiooni-teadlikkuse saavutamist. Muuhulgas arvasid eksperdid, et mehitamata platvormide treenimine jõukasutusreeglite rakendamiseks ei ole lõppkasutajatele jõukohane, eriti arvestades õiguslike reservatsioonide mitmekesisust riikidevahelises koalitsioonis. Tundmatu suurus on ka vastaspoolele kuuluv tehisintellektil baseeruva relvastuse operatsiooniline võimekus. Ekspertide hinnangul tuleb autonoomsete süsteemide kasutamisel silmas pidada ka algoritmilist rivaa-litsemist ehk lisaks oma üksuste efektiivsuse suurendamisele tuleb alistada ka vastase tehisintellekt. Nimetatud imperatiivid nõuavad kaitsevält koostöös liitlastega autonoomsete süsteemide integreerimist multidomeenilistes ühisoperatsioonides.

Uurimisküsimus 4: Millised tavakodanike ja ekspertide arvamused autonoomsete relvasüsteemide kohta ühtivad ? (**Uuring II–III**)

Mõlemas uuringus ühtivad arvamused kesketes teemades nagu inimese ja masinlike otsuste erinevused, vastutus autonoomsete relvade kasutamise tagajärgede eest ja nende arendamine või keelustamine teatud juhtudel. Uuringus osalejate jaoks on sõjalise jõu üle otsustamise delegeerimine masinale vastuvõetamatu. Kuigi mehitamata platvormide põhjustatud õigusrikkumiste menetlemine tekitab vaidlusi, ei tohi lubada vaikimisi tekkida lünka õigusemõistmises. Uuringutes ilmses konsensuslik seisukoht, et tehisintellekti kaasabil tehtud sõjalised otsused peavad olema inimestele mõistetavad. Seejuures ei tekkinud üksmeelt, kas toetada autonoomia täielikku keelustamist või liikuda arendamisega sammhaaval edasi, rakendades ranged regulatsioone ja kontrollimehhanisme. Erinevate sihtgruppide üksmeel uuringu kesketes teemades võib peegeldada seda, et neil on sarnased väärtused fundamentaalsetes küsimustes. Samas ei ole osalejatel (sh ekspertidel) väljakujunenud teoreetilist mõistmist ega äratuntavalt erinevaid paradigmasid seoses tehisintellekti sõjaliste rakendustega. Kuigi tegemist on murrangulise tehnoloogiaga sõjanduses, viitab see tähelepanek ka kontseptuaalse lähenemise puudumisele kaitsevält. Teisisõnu, lõppkasutajad ei taju ega mõtesta mehitamata sõjapidamist märkimisväärselt teisiti kui tavainimesed fookusgruppides. Nii võib öelda, et eelteadmiste puudumine tehisintellekti relvastamisest ei takista oluliselt diskursuses osalemist. Vaadates sotsioloogilisest aspektist on erinevate inimeste diskussioonis osalemine äärmiselt oluline, kuna temaatika puudutab suurt osa ühiskonnast.

Väitekirja kokku võttes on ilmne, et autonoomsete relvasüsteemidega kaasnevad unikaalsed väljakutsed. Pelgalt matemaatilisel arvutusvõimsusel ja eeltreenitud andmetel põhinev ratsionaalne väljund ei tohiks kunagi olla kaalukas osa militaarsetes otsustusprotsessides, kuna algoritmid ei suuda moraalseid tõeaktsidamisi genereerida. Siiski on ekslik mõelda sõjalisest tehisintellektist kui tapjarobotist. Pigem tuleb tuvastada autonoomse relvasüsteemi tugevused ja nõrkused ning

need oskuslikult ära kasutada. Mehitamata platvormide puhul on militaarvaldkonna jaoks hädavajalik tagada nende töökindlus ja läbipaistvus. On oluline, et lõppkasutajad mõistavad süsteemide toimimismehhanisme ning on teadlikud võimalikest manipulatsioonidest. Selleks tuleb relvajõudude liikmeid kaasata arendusprotsessi kõige varasemas staadiumis. Veelgi enam, osalejad rõhutasid (**Uuring II–III**), et huvigrupid nagu poliitikakujundajad, kaitsetööstussektor ja teadusasutused peavad viivitamatult jõudma konsensuseni, kuidas välja töötada ja jõustada regulatiivset raamistikku. Kahtlemata võivad tehisintellektiga varustatud süsteemid toimida sõjalise jõu olulise kordistajana, kuid selle potentsiaali ärakasutamise juures on vaja konsensust, et autonoomsed relvasüsteemid oleksid lahingutegevuses eetiliselt aktsepteeritud ja õiguslikult sanktsioneeritud. See on ja saab olema keerukas dialoog sõjanduse, diplomaatia, õigusloome, eetika ja ühiskondliku kokkulepe vahel.

Käesoleva teadustööga soovin teadvustada sõjalise tehisintellekti relvastamise ühiskondlikku tähtsust ja panustada diskussiooni sõjalise jõu delegeerimisel inimeselt masinale. Tehisintellekti relvastamises on väga palju teadmatust, mille ulatus ja mõju on suures osas veel uurimata. Sarnasel teemal erinevates kultuuri-ruumides tehtud uurimistööd oleksid eluliselt tähtsad, et ennetada ja maandada võimalikke riske, mis kaasnevad algoritmilise sõjapidamisega. Seetõttu pooldan jätkuvat vajadust hinnata algoritmide mõju sõjaliste juhtide otsustusprotsessidele – mil määral on otsustajad valmis oma hinnanguid muutma. Uurida tuleb ka inimese ja masina integreerimist nõudvat väljaõpet, et sõdurid (reservväelased) tuleks võimalikult efektiivselt toime autonoomsetele süsteemidega koos opereerides. Viimaks, kuigi mehitamata platvormide võidurelvastumine kogub hoogu, on algoritmilise sõjapidamise tulevik endiselt inimeste endi määrata. Soovin lõpetada tehisintellektist tulenevaid ohtusid uurinud pioneeri mõttega: „...kuna meil ei ole viise arvutite targaks tegemiseks, siis me ei tohiks anda arvutitele ülesandeid, mis nõuavad tarkust“ (Weizenbaum, 1976).

PUBLICATIONS

CURRICULUM VITAE

Name: Janar Pekarev
Date of birth: 14.07.1980
Phone: +372 56982 170
E-mail: janar.pekarev@ut.ee

Education:

2020–2024 University of Tartu, PhD, Sociology
2010–2012 Estonian Military Academy, MA, Military Science and Technology
2007–2014 University of Tartu, BA, Law
2003–2006 Estonian Military Academy, BA, Military Leadership

Professional experience:

2020–2024 Estonian Military Academy, Junior Researcher
2017–2020 Estonian Military Academy, Staff Officer
2016–2017 Joint Headquarters of the Estonian Defence Forces, Legal Adviser
2007–2013 Estonian Defence Forces, Officer-Instructor

Teaching experiences:

2024– University of Tartu, Military and Crises Sociology
2007–2023 Estonian Military Academy, MDMP, LOAC

Participation in research projects:

2020–2022 Ethical, social and legal aspects of unmanned weapon systems (R-012)
2020–2020 Mapping and modelling the issues related to product development in the defence industry of a small state in the context of global value chains to enhance the competitiveness of the defence industry (R-010)
2018–2022 Human resource survey in Estonian Defence Forces (R-006)

Selected publications:

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C., Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society*.

Wagner, W., Viidalepp, A., Idoiaga-Mondragon, N., Talves, K., Lillemäe, E., Pekarev, J., Otsus, M. (2023). Lay representations of artificial intelligence and autonomous military machines. *Public Understanding of Science*, 32 (7), 926–943.

- Pekarev, J. (2023). Dilemmas of representation of autonomous weapon systems (AWS). A focus group perspective and in-depth interviews with military personnel. EuroISME Annual Conference 2023: The Future of Military Ethics, The International Society for Military Ethics in Europe.
- Pekarev, J. (2022). Moral Guidance for Autonomous Weapons Systems (AWS). A focus group perspective. Conference of the International Society of Military Sciences: Promoting Peace and Security in a new incomprehensible and non-linear world, Military University Institute.
- Pekarev, J. (2021). Towards Trustworthy Autonomous Weapon Systems: A compendium of ethical aspects. 16th Biennial Conference of ERGOMAS: Dynamic armed forces in changing societies: challenges for research in and on the military.

ELULOOKIRJELDUS

Nimi: Janar Pekarev
Sünniaeg: 14.07.1980
Telefon: +372 5698 2170
E-post: janar.pekarev@ut.ee

Haridus:

2020–2024 Tartu Ülikool, sotsioloogia doktoriõpe
2010–2012 Kaitseväe Akadeemia, sõjateadus- ja tehnoloogia, magistriõpe
2007–2014 Tartu Ülikool, õigusteadus, bakalaureuseõpe
2003–2006 Kaitseväe Akadeemia, sõjaväeline juhtimine, diplomiõpe

Teenistuskäik:

2020–2024 Kaitseväe akadeemia, nooremteadur
2017–2020 Kaitseväe akadeemia, staabiohvitser
2016–2017 Kaitseväe peastaap, jurist
2007–2013 Kaitsevägi, ohvitser-instruktor

Õpetamiskogemus:

2024– Tartu Ülikool, militaar- ja kriisisotsioloogia
2007–2023 Kaitseväe Akadeemia, otsuse vastuvõtmise protsess, relvakonfliktioigus

Osalemine uurimisprojektides:

2020–2022 Mehitamata relvasüsteemide eetilised, sotsiaalsed ja õiguslikud aspektid (R-012)
2020–2020 Väikeriigi kaitsetööstuse tootearenduse kitsaskohtade kaardistamine ja mudeldamine globaalsete väärtusahelate kontekstis parema konkurentsivõime tagamiseks (R-010)
2018–2022 Riigikaitse alase inimressursi kompleksuuring Kaitseväes (R-006)

Valitud publikatsioonid:

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C., Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society*.

Wagner, W., Viidalepp, A., Idoiaga-Mondragon, N., Talves, K., Lillemäe, E., Pekarev, J., Otsus, M. (2023). Lay representations of artificial intelligence and autonomous military machines. *Public Understanding of Science*, 32 (7), 926–943.

- Pekarev, J. (2023). Dilemmas of representation of autonomous weapon systems (AWS). A focus group perspective and in-depth interviews with military personnel. *EuroISME Annual Conference 2023: The Future of Military Ethics, The International Society for Military Ethics in Europe*.
- Pekarev, J. (2022). Moral Guidance for Autonomous Weapons Systems (AWS). A focus group perspective. *Conference of the International Society of Military Sciences: Promoting Peace and Security in a new incomprehensible and non-linear world, Military University Institute*.
- Pekarev, J. (2021). Towards Trustworthy Autonomous Weapon Systems: A compendium of ethical aspects. *16th Biennial Conference of ERGOMAS: Dynamic armed forces in changing societies: challenges for research in and on the military*.

DISSERTATIONES SOCIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Veronika Kalmus.** School textbooks in the field of socialisation. Tartu, 2003, 206 p.
2. **Kairi Kõlves.** Estonians' and Russian minority's suicides and suicide risk factors: studies on aggregate and individual level. Tartu, 2004, 111 p.
3. **Kairi Kasearu.** Structural changes or individual preferences? A study of unmarried cohabitation in Estonia. Tartu, 2010, 126 p.
4. **Avo Trumm.** Poverty in the context of societal transitions in Estonia. Tartu, 2011, 215 p.
5. **Kadri Koreinik.** Language ideologies in the contemporary Estonian public discourse: With a focus on South Estonian. Tartu, 2011, 128 p.
6. **Marre Karu.** Fathers and parental leave: slow steps towards dual earner/dual carer family model in Estonia. Tartu, 2011, 125 p.
7. **Algi Samm.** The relationship between perceived poor family communication and suicidal ideation among adolescents in Estonia. Tartu, 2012, 121 p.
8. **Tatjana Kiilo.** Promoting teachers' efficacy through social constructivist language learning: challenges of accommodating structure and agency. The case of Russian-speaking teachers in Estonia. Tartu, 2013, 156 p.
9. **Ave Roots.** Occupational and income mobility during post-socialist transformation of 1991–2004 in Estonia. Tartu, 2013, 130 p.
10. **Tarmo Strenze.** Intelligence and socioeconomic success A study of correlations, causes and consequences. Tartu, 2015, 119 p.
11. **Mervi Raudsaar.** Developments of social entrepreneurship in Estonia. Tartu, 2016, 141 p.
12. **Ero Liivik.** Otsedemokraatia Eestis: õigussotsioloogilisi aspekte. Tartu, 2017, 166 p.
13. **Mai Beilmann.** Social Capital and Individualism – Collectivism at the Individual Level. Tartu, 2017, 145 p.
14. **Rainer Reile.** Self-rated health: assessment, social variance and association with mortality. Tartu, 2017, 123 p.
15. **Katri Lamesoo.** Social Construction of Sexual Harassment in the Post-Soviet Context on the Example of Estonian Nurses. Tartu, 2017, 185 p.
16. **Andu Rämmer.** Sotsiaalse tunnetuse muutused Eesti siirdeühiskonna kontekstis. Tartu, 2017, 230 p.
17. **Kadri Rootalu.** Antecedents and consequences of divorce in Estonia from longitudinal and multigenerational perspectives. Tartu, 2017, 128 p.
18. **Kairi Talves.** The dynamics of gender representations in the context of Estonian social transformations. Tartu, 2018, 129 p.
19. **Aare Kasemets.** Institutionalisation of Knowledge-Based Policy Design and Better Regulation Principles in Estonian Draft Legislation. Tartu, 2018, 252 p.

20. **Dagmar Narusson.** Personal-recovery and agency-enhancing client work in the field of mental health and social rehabilitation: Perspectives of persons with lived experience and specialists. Tartu, 2019, 139 p.
21. **Oliver Nahkur.** Measurement of Interpersonal Destructiveness: the Societal Perspective. Tartu, 2019, 164 p.
22. **Tayfun Kasapoglu.** Algorithmic Imaginaries of Syrian Refugees: Exploring Hierarchical Data Relations from the Perspective of Refugees. Tartu, 2021, 152 p.
23. **Kristjan Kikerpill.** Crime-as-communication: detecting diagnostically useful information from the content and context of social engineering attacks. Tartu, 2021, 162 p.
24. **Taavi Laanepere.** Looking at the Military Service Readiness of Estonian Reserve Soldiers through the Prism of Bourdieu's Theory of Practice. Tartu, 2021, 174 p.
25. **Tiia-Triin Truusa.** The entangled gap: the male Estonian citizen and the interconnections between civilian and military spheres in society. Tartu, 2021, 159 p.
26. **Kadri Soo.** School as a source of child subjective well-being in the framework of children's rights: Perspectives of children and young adults. Tartu, 2023, 146 p.
27. **Hans Hõrak.** Development of a computer vision-based privacy-preserving automatic observation method for measuring physical activity in school. Tartu, 2023, 158 p.
28. **Märt Masso.** Employment relations in Estonia: employee control, participation and work accommodation in co-determining working conditions. Tartu, 2023, 133 p.
29. **Simone Eelmaa.** The social categorization of sexual abuse. Tartu, 2024, 247 p.
30. **Kristi Paron.** Child-patient autonomy: interplay between normativity and relationality. Tartu, 2024, 118 p.