FACULTY OF LAW
Lund University

Luisa Teresa Hedler Ferreira

Driverless Cars and Merciless Robots – a comparison between the legal frameworks of autonomous driving and weapon systems.

JAMM07 Master Thesis

International Human Rights Law
30 higher education credits

Supervisor: Gregor Noll

Term: Spring 2018

Abstract

Artificial Intelligence (AI) technology advances quickly through society, with driverless cars presented as a revolutionary technological advance that can make roads safer and transport more efficient, whereas automated weapons are either viewed as undesirable, or as an unavoidable development. The objective of the present work is to, through the comparison between automated driving and automated warfare, identify the challenges in articulating legal responsibility with the usage of AI in warfare. This comparison is possible because they face similar dilemmas in their interactions with the legal world – especially in the distribution of responsibility in cases where there is targeting involved. Through a brief explanation of different scales of automation, it is possible to clarify that artificial intelligence can be applied in different parts the decision-making process, and with different degrees of autonomy from human subject – and that most of the current focus, when it comes to automated weapons, is in problematizing the possibility of decision-making in the machine, insisting in the reinforcement of a 'relevant human control'. When it comes to driverless cars, however, decision-making in cases of car accidents is seen as a given, which makes the legal community discuss further, delving into earlier stages of decision-making, such as data processing and selection. This comparison brings to light a few shortcomings of the ability of legal systems in general to deal with arrangements in which autonomy and responsibility do not report directly to a human agent.

Keywords: Artificial Intelligence, Humanitarian Law, Driverless Cars, Automated weapons.

Contents

Abbreviations

ADS – Automated Driving Systems

AI – Artificial Intelligence.

CCW – Convention on Certain Conventional Weapons

HRW – Human Rights Watch.

IHL – International Humanitarian Law

IHRL – International Human Rights Law

LAWS – lethal autonomous weapon systems

SAE – Society of Automotive Engineering

UN – United Nations

USDoD – United States Department of Defense

1. Introduction

The creations of humanity surpass us. They surpass us in speed, in strength, in lethality – and now, increasingly, in our capacity to think[1]. The employment of machines with Artificial Intelligence (AI) is growing in so many fields because it has the potential to make up for the areas of the brain where we are slow and imprecise – but with these new possibilities come questions and dilemmas that are already hard choices and delicate balances for human beings. The challenge, then, is how to program machines that are capable of making those decisions in a satisfactory way.

The ability to act according to laws is one of the many challenges that AI faces. After all, human laws are quite unlike simple commands that can be easily programmed – there is plenty of space for interpretation, for balancing principles, inexact parameters that were built into the law precisely because of the imprecision of human beings. They are fruit of political processes that do not necessarily conform to a systematic logic, and are subject to many different contextual and political pressures. Not to mention how complex the interaction between different systems – international and domestic, criminal and civil, humanitarian and human rights – can be.

This very human complexity in the interpretation and application of law is very visible in Humanitarian Law – a part of International Law that sits somewhat uncomfortably in the current International Law system, especially since the existence of Humanitarian Law in itself carries the paradox of regulating a type of human conduct – *warfare* – that is currently forbidden by the United Nations Charter, and the avoidance of which is one of the main self-proclaimed objectives of the foundation of the current form of multilateral relations.

As complex as Humanitarian Law already is, the advances in the technology of AI challenge structures and workings of this particular legal system in new ways – how these new weapons can be classified, how their actions can fit the existing legal categories, and to which degree current legislation is equipped to deal with the

[1] Eva Zackova, 'Intelligence Explosion Quest for Humankind', in Jan Romportl, Eva Zackova, and Jozef Kelemen (eds.), *Beyond Artificial Intelligence - the Disappearing Human-Machine Divide* (New York: Springer, 2015), 31-44.

diminishing role, or even removal, of a natural person from the decision-making process, a change that can lead to severe material (and legal) repercussions.

Moreover, the way laws are made can also influence the development and shaping of these technologies – as a factor that new actors in warfare must adapt to. The development of new technologies can be done either to comply or to circumnavigate Humanitarian Law, designed to navigate the existing legal body in an instrumental way to maximise their advantage in combat without disturbing the formal restraints dictated by the law.

Autonomous weapons are a big cause of alarm to the international community, with outcries to ban 'killer robots' mixed between claims of a third technological revolution in warfare and a veiled competition between countries about it. Between these voices, there are some who claim that automated units could be an asset in war, and perhaps capable of being even more humane than humans themselves.

When it comes to analysing the interaction between laws and autonomous weapons, it is still difficult to get extensive information – be it for the slowness of the pace with which International Law develops, or be it by the secretive nature of military development. However, there are other applications of AI outside of the military in which discussions are being held about how AI will adapt to legal systems – and how legal systems will adapt to the new technologies. The rapid development of automated driving technologies (also called driverless cars) are being openly discussed, both in a domestic setting and on an international level. Advertised as the future of transit, leading to safer, more accessible human mobility, some levels of automation are already available for commercialisation, while driverless models are slowly being tested in public streets.

Even though automated driving and automated warfare might mobilize different areas of law, and cover activities that are very different, they present many common characteristics that can make a comparative analysis quite useful.

Chapter 2 begins with a general description of the concept of AI, and the introduction of different scales of automation that measure the levels in which the main functions of a given technology can be operated without human input. After that, a brief overview of the current legal context of each field is made.

The comparison between automated driving and weapons systems is made in chapter 3 by introducing the main dilemmas that both of these technologies have in common, and analysing the way each field approaches them.

Therefore, the objective of the present work is to, through the comparison between automated driving and automated warfare, identify the challenges in articulating legal responsibility with the usage of AI in warfare, and answer the following research question: how do the car industry and the military industry articulate legal responsibility for its uses of AI?

2. Artificial Intelligence – developments and legislation

The first step in comparing automated cars and automated weapons is to take a cursory look into the technology that made both of them possible. Artificial intelligence, as the name suggests, is the simulation of different aspects of human intelligence by machines[2]. When it began to emerge as a field of studies in the 50s, the original intention was to simulate the rationality and intelligence of humans, but it soon divided itself into simulation of different aspects of human intelligence, adapted to specific tasks[3] - be it perception, communication, speech processing, pattern recognition, orientations, etc. Rather than simulating an entire human being through machines, most AI developers are more focused on developing machines for problem-solving in specific tasks[4]. Therefore, the first point that needs to be made clear while discussing the application of AI in different fields is that there is no monolithic AI that can be used in all contexts – the programming of each device can simulate different areas of human intelligence, be adapted for different tasks and have vastly different characteristics.

In this sense, AI is an umbrella term, used to describe an enormous variety of different machines, with differences in programming and level of complexity. The machines can range from game-playing software, which have been only fed the rules of a game as an input, to "Artificial General Intelligence" projects, who seek to emulate the human mind as a whole[5].

In this sense, a difference is made between what is called weak and strong AI. The first one refers to narrower uses of AI, which only emulate one or a few specific aspects of human intelligence, whereas the second demands the capacity to perform broader, more complex tasks that require a more complex and multifaceted type of intelligence[6].In the present work, however, the focus will be in two specific uses of AI: automated driving systems (ADS), and lethal autonomous weapon systems (LAWS).

---

[2] Ibid.
[3] Ibid.
[4] Ibid., p.32-33
[5] Ibid.
[6] Michael Carl Haas and Sophie-Charlotte Fischer, 'The Evolution of Targetet Killing Practises: Autonomous Weapons, Future Conflict, and the International Order', *Contemporary Security Policy,* 38/2 (2017), 281-306., p. 287

The multitude of functions that can be automated leads to a situation where there is not one single way of incorporating AI into weapons or cars alike – there are different processes that can be transferred from humans to machines, as well as levels of complexity in which this can be done. In this sense, it is important to clarify the different types of AI that can be used in either field – as well as the models used to explain them.

2.1. Levels and functions of automation

Whereas AI, in the present work, is examined in its potential to enable the automation of both cars and weapons, the two concepts are not synonymous. Automation, according to the Merriam-Webster dictionary, can be defined as an *"automatically controlled operation of an apparatus, process or system by mechanical or electronic devices that take the place of human labor*"[7]. In this sense, not every automated process requires AI to work. Technology as crude as landmines, for example, can be counted as an example where there is automation without the use of particularly sophisticated thinking processes.

Different levels of automation, therefore, describe different relationships that distribute the decision making between the human being which is the user, and the system. The traditional model for measuring levels of automation was refined in 2000, and presents a scale that varies from total human control of decisions (level 0) to a scenario where the system ignores the human user completely, not even providing information about the decision taken (level 10)[8]:

10. The computer decides everything, acts autonomously, ignoring the human.
9. informs the human only if it, the computer, decides to
8. informs the human only if asked
7. executes automatically, then necessarily informs the human, and
6. allows the human a restricted time to veto before automatic execution, or
5. executes that suggestion if the human approves, or
4. suggests one alternative
3. narrows the selection down to a few, or

[7] https://www.merriam-webster.com/dictionary/automation
[8] Dale Richards and Alex Stedmon, 'To Delegate or Not to Delegate: A Review of Control Frameworks for Autonomous Cars', *Applied Ergonomics,* 53 (2016), 383-88., p.385.

2. The computer offers a complete set of decision/action alternatives, or

1. The computer offers no assistance: human must take all decisions and actions.[9]

Aside from creating a model with different levels of automation, Parasuraman et al. (2000) also created a simplified four-stage model of human information processing, separating the different operations to which the automation would apply. These four different functions to which automation can be applied are:

1. information acquisition;
2. information analysis;
3. decision and action selection;
4. action implementation[10].

Automation in information acquisition means that the sensing and registration of input data is left to the machine, instead of an outside source, programme or operator having to manually feed information into the system. At a lower level, this might mean that a system moves sensors around, or "locks" into detected targets, but at a higher level the raw data acquired can be sorted into priorities and organised. Examples of this type of automation are radars that are programmed to 'follow' targets[11].

When it comes to information analysis, the automation process can substitute human working memory and some logical processes, for example, by utilising raw data to make projections, or even information integration, where a multitude of variables can be combined into a result, managing information so the human agent is presented with one or a few choices[12]. One example of automation in information analysis already in use is in the criminal system. In several jurisdictions across the United States, risk assessment algorithms are being used to determine the likelihood of a criminal to reoffend[13]. In this case, an assessment that was hitherto made by the judges, using their own discernment and processing all evidence and information available to them in the procedure, can now

[9] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens, 'A Model for Types and Levels of Human Interaction with Automation', *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans,* 30/3 (2000), 286-97., p.287.
[10] Ibid., p. 288.
[11] Ibid., p. 288.
[12] Ibid., p.288.
[13] Keith Kirkpatrick, 'It's Not the Alrorithm, It's the Data - in Risk Assessment and Predictive Policing, Biased Data Can Yield Biased Results.', *Communications of the Association for Computing Machinery,* 60/2 (2017), 21-23.

be delegated to an automated system, which in turn can influence the length of sentencing[14].

After processing information, the next step that can be automated is the decision and action selection – that is, choosing between alternatives and courses of action. In this sense, the machine replaces human selection between alternatives with machine decision making. In aviation, for example, a software can automatically plan routes as to avoid bad weather[15]. This stage of decision-making is the one that most comfortably fits the model of levels of automation – in which the decision made by the system can be presented to the user as a suggestion, in the lower levels of automation, or just be enforced without even the knowledge of the user, as it is in higher levels.

Finally, the final stage of decision-making is the automation of action – in other words, of machines executing the response to the data collected, analysed and course of action selected. Their variation is on the amount of activity that is performed by the machine, versus how much is done manually[16].

This separation is very helpful to give more precision to the debate on automation – and to ascertain which kinds of operation are potentially more problematic in terms of legal responsibility. A single machine could present different levels of automation in each dimension, and also vary the levels of automation in each level according to circumstances or the will of the user[17]. A computer keyboard with an auto-correct feature is an example of simple action automation – parts of the action of typing are suggested – or even automatically substituted – by the software.

Considering the almost infinite different possibilities of uses of AI, both in the automotive and warfare industries, there are multiple forms of using them, as well as multiple levels of automation that can be reached by the employment of such technology. Of course, the stages of automation depend highly on which functions are expected to be executed by a particular machine – and therefore, the measurement for automation in cars and in warfare follow slightly different standards.

[14] Alyssa M. Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review,* 103/303 (2017), 304-29., p. 305.
[15] Parasuraman, Thomas B. Sheridan, and Wickens, 'A Model for Types and Levels of Human Interaction with Automation', (, p. 289.
[16] Ibid., p. 289.
[17] Ibid., p. 289.

2.1.1. Automotive industry: five stages of automation

The Society of Automotive Engineering (SAE) has issued a recommended practise with the taxonomy and definitions regarding automated vehicles (J3016-201609)[18], which classifies levels of automation based on the aptitude of an automated system to perform dynamic driving tasks[19]. The levels of automation range from 0 to 5, and can be summarised in the following way:

Level 0 – There is no automation, and the human driver has total control at all times.

Level 1 – There is some form of driver assistance, which might steer, accelerate or decelerate, and expects the human driver to perform all other tasks and have their attention on the road always.

Level 2 – There is partial automation, where the driver can be relieved of at least two primary control functions. In practise, that means the human driver can have hands off the wheel, and the foot off the pedal. However, the driver still has to monitor the driving process, and be ready to take over at any moment.

Level 3 – There is conditional automation. The driver can cede full control under certain conditions, and the vehicle will monitor when those conditions change and hand control back to the driver. In this phase, the driver is only required to have occasional control, when the vehicle deems the situation too unpredictable to operate safely.

Level 4 – There is a high level of automation. The driver is not expected to engage with the vehicle except to enter the directions. However, the vehicle might only operate in certain environment or under certain conditions. Once the conditions are met, the car would be highly automated.

Level 5 – There is full automation. The vehicle can perform as well as a human driver, under all types of condition, and there is no need of a human being to supervise the vehicle in any setting.

The measure of this scale is how little the driver has to do in terms of the main functions of the vehicle, and the relationship that an individual human agent – the driver – has

---

[18] Society of Automotive Engineering, 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles', in Sae (ed.), (J3016_201609, 2017).

[19] That is, the main tasks that must be performed for a car to work. ABS breaks and other emergency settings which interfere only occasionally and punctually in the driving are not included in this classification.

with the system. There are many nuances in the intermediary stages between stage 0, where the driver has full control and responsibility, and stages 4 and 5, where all humans in the car can be regarded as passengers.

The critical functions of driving – accelerating, breaking, following directions, changing lanes, avoiding obstacles, etc – is a highly complex task that demands a complex engagement of the human mind. In this sense, the task of driving demands information collection, processing, selection of choices and action – the full spectrum of the four functions of automation in the model presented by Parasurama et al, (2000).

Therefore, in cars that have Level 1 or 2 of automation in the SAE scale, there is a limited automation in the *action* function of driving-related tasks – a few of the actions of driving are delegated to the machine, even if the human capacity for decision-making and acting cannot be fully disengaged while the vehicle is operating.

A shift in the machine-driver distribution of tasks occurs in Level 3: instead of the driver maintaining situational awareness (SA) at all times, choosing when to deploy the automated system, now the system decides when to give back control to the driver[20]. There is, therefore, a higher level of automation in the decision-making process of delegation.

In levels 4 and 5, the function of action is completely taken over by the system. The only difference between them is that level 4 systems still have space for some input of the driver in the function of choice, whereas level 5 completely alienates the user.

2.1.2. Autonomous weapons: "In-the-loop" and "out of the loop"

While the measure for automation in cars is the capacity to perform the main operations of a vehicle without the assistance of a human driver, in a similar fashion the level of autonomy of LAWS are measured by their capacity to select and engage targets.

The United States Department of Defense (USDoD), in their directive concerning autonomy in weapon systems (DoDD 3000.09), differentiates between autonomous and semi-autonomous weapon systems in the following way:

[20] Except, of course, in cases where the driver decides to override the system, which is a possibility that is implicit in all levels of automation in the SAE scale except for level 5.

Semi-autonomous weapons engage individual targets or specific target groups only after being selected by human operators. They proceed do provide a number of examples of this application:

> Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement. "Fire and forget" or lock-on-after-launch homing munitions that rely on TTPs to maximize the probability that the only targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator. (US-DoD 2012: 14)[21].

Automated weapons, on the other hand, are defined by having the capacity to select and engage targets without the necessity of intervention by a human operator[22] – the main objectives of a mission are established by a human operator, but there is the possibility of operation without further input.

In their report about Autonomous Weapons and legal responsibility, Human Rights Watch (HRW) classified the different levels of automation according to their levels of autonomy[23] – in this case, meaning how much human involvement was in their actions. The division goes as follows:

> • Human-in-the-Loop Weapons: Robots that can select targets and deliver force only with a human command;
>
> • Human-on-the Loop Weapons: Robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions; and
>
> • Human-out-of-the Loop Weapons: Robots that are capable of selecting targets and delivering force without any human input or interaction.[24]

It is noticeable that this description is considerably less detailed than the SAE – in a rough comparison, the categories outlined by Human Rights Watch could only apply for level 4 and 5 automated vehicles. One of the factors that could explain the difference are the transitional needs of automotive technology. The same critical functions (accelerating, breaking, changing lanes), at a level 3, for example, can be performed by

---

21 Us-Dod, 'Directive on Autonomy in Weapons Systems', in Us Department Of Defense (ed.), (2012).
22 Ibid., p. 13
23 Harvard International Human Rights Clinic Human Rights Watch, 'Mind the Gap - the Lack of Accountability for Killer Robots', (2015).
24 Ibid., p. 6

either the machine or the human driver, whereas the technology of automated weapons are different than a robot-car that would merely substitute a human operator by a machine.

In terms of the functions that are considered relevant, both the USDoD and HRW focus heavily in the automation of actions: even in the 'lower' stages of autonomy, it is presupposed that the systems are already capable of gathering and process information, as well as make choices. In this specific case, the choice that has to be made is the selection of targets, and even in the more nuanced description of semi-automated weapons by the USDoD, even when the ultimate decision on selection of target might fall to the human agent, operations of pre-selection or cueing a human agent are described as examples. In any case, the main difference between semi-autonomous and autonomous weapons relies on who has the capacity to initiate the engagement. The same happens with the HRW definitions, where the level of control of engagement (the action function) is the limit between each category.

2.2. Between automation and autonomy.

After perusing the different levels of complexity that AI can reach in both vehicles and warfare, is important to notice that there is a significant difference in language related to the distinct fields: whereas the automotive industry mostly refers to *automated* vehicles, when it comes to weapons the adjective used is *autonomous*. As Ekelhof points out in the context of the discussion of autonomous weapons, the term 'autonomy' is heavy with multiple meanings and different philosophical implications, and can be interpreted in several ways[25].

Originally, the term 'autonomy' has a Greek origin, and translates directly to 'self-ruling', that is, the capacity of existing within their own norms and terms[26] - but when talking about AI, there is a multitude of meanings that can be ascribed to it, from a more mundane capability of operating without human supervision, to apocalyptic scenarios of sentient, self-governing robots who define their own agenda.

[25] Merel A. C. Ekelhof, 'Complications of a Common Language: Why Is It So Hard to Talk About Autonomous Weapons?', *Journal of Conflict & Security Law,* 22/2 (2017), 311-31., p. 323

[26] Hamid R. Ekbia, 'Heteronomous Human and Autonomous Agents: Towards Artificial Relational Intelligence', in Jan Romportl, Eva Zackova, and Jozef Kelemen (eds.), *Beyond Artificial Intelligence - the Disappearing Human-Machine Divide* (Switzerland: Springer, 2015)., p. 67.

Firstly, one common use of the term 'autonomy' is as a synonym to 'automation' – referring, therefore, to the capacity of the system to perform without human input[27]. This is the definition that has been used hitherto – while describing and comparing systems of automation, the only aspect that has been considered so far is the distributions of agency between machines and humans, a sense in which automation and autonomy can be used as equivalents. When it comes to automated driving, in many sources, the terms 'automation' and 'autonomous' have been used interchangeably[28].

However, that definition is inadequate when considering examples of weapons systems that can be utilized and hit a target independently of the presence or supervision of their owners. Dumb landmines or a tripwire sentry gun, for example, are independent of human interference to be deployed – therefore *automated* – but do not have the sophistication to select a target, and merely react to a simple trigger[29].

A slightly more complex concept of autonomy, as described by Haas and Fischer (2017)[30], focuses on the command and control relationship between humans and machines – a notion that coincides with the USDoD and HRW classifications of autonomous weapons.

In this sense, it is easier to understand the conflation between automation and autonomy in the case of automated driving systems. Whereas a landmine – automated but not autonomous – is capable of targeting (albeit at random) and engaging against that target without direct control or instructions of the entity that deployed it, such a crude technology could not possibly accomplish the task of driving. Navigating in public roads, with the complexity of the situations in streets with other vehicles, pedestrians, weather conditions and so many other factors demands the type of decision making functions that go beyond mere automation, demanding a deeper level of *autonomy* from the system[31], now signified to mean something more than automation.

[27] Ekelhof, 'Complications of a Common Language: Why Is It So Hard to Talk About Autonomous Weapons?', (, p. 324.
[28] Cesare Bartolini, Tamás Tettamanti, and István Varga, 'Critical Features of Autonomous Road Transport from the Perspective of Technological Regulation and Law', *Transportation Research Procedia,* 27 (2017), 791-98.
[29] Rebecca Crootof, 'Autonomous Weapons Systems and the Limits of Analogy', (Yale Law School, 2018)., p. 7.
[30] Haas and Fischer, 'The Evolution of Targetet Killing Practises: Autonomous Weapons, Future Conflict, and the International Order', (, p. 285.
[31] Richards and Stedmon, 'To Delegate or Not to Delegate: A Review of Control Frameworks for Autonomous Cars', (, p. 384.

.

These definitions however, do not take into account the level of complexity of the decisions that a system has to make, nor does it differentiate between the origin and the quality of the decisions being made by the system. So when it comes to the level of sophistication of autonomy, Boulanin (2016) uses the division between automatic, automated and autonomous systems[32]. While automatic machines respond to fixed commands, mechanic sensory inputs, and are not capable of varying their behaviour in the context (such as the much mentioned dumb landmine), the other two categories would encompass machines and systems that could respond to variations in the environment – and the varying levels of complexity would determine whether it was just automated or autonomous. Haas and Fischer (2017) consider that missile defence systems, which are already in use in many places of the world, could be considered highly automated, but not autonomous, since they provide only predictable outcomes based on pre-programmed rule-based systems[33]. In this sense, autonomy could only be said to be present when systems were apt to perform tasks and make decisions based on parameters *outside* of what they have been programmed to do[34].

The possibility of a system acting outside of their original programming is tied to the idea of "machine learning" – which is heralded as a possible solution for machines that operate in highly complex and volatile environments, such as public transit or battlefields[35]. Machine learning is when systems are programmed to adapt to new experiences and incorporate them into their future actions, without the need of further programming from a human. There is a subset of machine learning called "deep learning", in which AI structures mimic the human brain in order to allow for this adaptation to the environment[36].

A third form of assessing autonomy in systems – and which is more applicable to the current work – is defining autonomy based on which functions of a system are being made autonomous – therefore, a functional approach to autonomy, that assesses the

[32] Vincent Boulanin, 'Mapping the Development of Autonomy in Weapon Systems - a Primer on Autonomy', *Stockholm International Peace Research Institute Working Paper* (2016)., p. 3.
[33] Haas and Fischer, 'The Evolution of Targetet Killing Practises: Autonomous Weapons, Future Conflict, and the International Order', (, p. 285.
[34] Ekelhof, 'Complications of a Common Language: Why Is It So Hard to Talk About Autonomous Weapons?', (, p. 324-325.
[35] Haas and Fischer, 'The Evolution of Targetet Killing Practises: Autonomous Weapons, Future Conflict, and the International Order', (, p. 282
[36] Ibid.

human-machine interface in the different operations that might be required of each system[37].

Whereas the issue of autonomy might be of great importance in definitional or philosophical discussions, it is important to keep in mind that it is precisely the capacity of operating with limited or no human input that introduces legal complexities in the employment of AI – in a legal system made to recognise and accommodate human agents, the possibility of actions being autonomously performed by non-human actors poses difficulties when it comes to assigning responsibility.

2.3. Autonomous Driving Systems

Considering the SAE scale of automation, the automotive industry has not, as of yet, gone beyond the large-scale commercialisation of partially automated vehicles (level 2), but there is a strong trend towards the development of more and more autonomous cars, to the point where even the most conservatives estimates give only 10 to 15 years until such technology is available to the public[38].

The main justification for the stance of progressively alienating the human element in the act of driving stems from a concern with public safety. Not only the car developing companies[39], but researchers[40], governments[41] and even sectors of the UN[42] hail driverless cars as a life-saving technology. It is expected, after all, to greatly diminish the number of accidents on the roads, and promote a safer, more efficient, more inclusive and environmentally friendly way of transportation[43].

In terms of International Law, there are two main instruments which seek to harmonize domestic rules involving automotive traffic: The Geneva Convention on Road Traffic (1949) and the newer Vienna Convention on Road Traffic (1968) (hereby mentioned as

[37] Boulanin, 'Mapping the Development of Autonomy in Weapon Systems - a Primer on Autonomy'., p. 4.
[38] Tract Hresco Pearl, 'Fast & Furious: The Misregulation of Driverless Cars', *NYU Ann. Surv. Am. L.,* 73/19 (2017) at 25.
[39] Tesla, 'An Update on Last Week's Accident', (2018a).
[40] Robert Sparrow and Mark Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', *Transportation Research Part C,* 80 (2017), 206-15.
[41] Federal Ministry of Transport and Digital Infrastructure Of Germany, 'Ethics Comission on Automated and Connected Driving - Report of June 2017.', (2017).
[42] Unece, 'Unece Paves the Way for Automated Driving by Updating Un International Convention', (2016).
[43] Heather Bradshaw-Martin and Catherine Easton, 'Autonomous or 'Driverless' Cars and Disability: A Legal and Ethical Analysis.', *European Journal of Current Legal Issues,* 20/3 (2014).

the Vienna Convention)[44]. The enduring relevance of the former is due to the fact that the United States and the United Kingdom have not ratified the latter. Both of them, however, initially define driver as *a person,* and have the requirement that there should be a driver in control of the vehicle (or herding animals) at all times. In the wording of the Vienna Convention:

> Article 1: Definitions, Section (v)
>
> "Driver" means any person who drives a motor vehicle or other vehicle (including a cycle), or who guides cattle…
>
> (…)
>
> Article 8 – Drivers.
>
> 1 Every vehicle or moving combination of vehicles shall have a driver.
>
> 3.Every driver shall possess the necessary physical and mental ability and be in a fit physical and mental condition to drive.
>
> (…)
>
> 5.Every driver shall at all times be able to control his vehicle or to guide his animals.
>
> (…)
>
> Article 13 – Speed and distance between vehicles.
>
> 1. Every driver of a vehicle shall in all circumstances have his vehicle under control so as to be able to exercise due and proper care and to be at all times in a position to perform all manoeuvres required of him[45].

The Vienna Convention, however, was recently amended to allow for the possibility of automated driving systems – with the condition that a human driver should have the possibility to take control and override the system[46].

Therefore, international regulation as it is today permits vehicular automation up to level 4 in the SAE scale. When applying this situation to the human-machine interface model of gauging autonomy, it is possible to infer that the current international regulations are able to accommodate semi-autonomous cars, where humans are on the

[44] Anonymous, 'Vienna Convention on Road Traffic (1968) U.N.E.A.S.Council', (1968).
[45] Ibid.
[46] The ammendment to article 8 reads as follows: 5bis: Vehicle systems which influence the way vehicles are driven shall be deemed to be in conformity with paragraph 8 of this article and with paragraph 1 or article 13, when they are in conformity with the conditions of construction, fitting and utilization according to international legal instruments concerning wheeled vehicles (…). Vehicle systems which influence the way vehicles are driven and are not in conformity with the aforementioned conditions of construction, fitting and utilization, shall be deemed to be in conformity with paragraph 5 of this Article and with paragraph 1 of article 13, when such systems can be overridden or switched off by the driver.

loop. One could argue that full automation is also supported – with the caveat that there must be at least the possibility of the system returning to the control of a driver. In this sense, even full automation is not forbidden by international regulations – on the condition that it might be reversible.

This shows a high level of acceptance of delegating critical functions to machines, but the last threshold – doing away with the figure of a human driver, even if only with the potential to take control – is not yet supported by international law. The relevance of this last barrier towards automation becomes clear when it comes to the mobility of people with disabilities, some of whom have no possibility of meeting the requirements to be drivers[47].

The Vienna Convention on Road Traffic sets the standard for 74 countries in the world, and Europe, the US and China have specific similar regulations, with strong trends towards developing laws to contemplate autonomous driving technologies[48]. The US, for example, has federal guidelines issued by their National Highway Traffic Safety Administration, but only as a voluntary guidance[49]; each state has jurisdiction to enact their own laws, according to their economic interest in the development of the automotive industry against public concerns about the possible dangers posed by the testing of progressively more autonomous cars in their public streets.

Germany, on the other hand, went a step further in tackling the legal and moral complexities of autonomous driving technology. The State has assembled a Committee of Ethics to report on the main issues that might arise from driverless car technology, from targeting decisions in case of accidents to the legal uses of all data produced by the car's sensors[50]. The works of the commission resulted in 20 ethical principles for driverless cars, which has been adopted by the German federal government[51].

The development of this new technology is not without setbacks and challenges, and every car accident involving automated driving systems – especially if they are lethal – tend to gather an enormous amount of scrutiny from society in general, the press and

---

[47] Heather Bradwhaw-Martin and Catherine Easton, 'Autonomous or 'Driverless' Cars and Disability: A Legal and Ethical Analysis', *European Journal of Current Legal Issues,* 20/3 (2014), 1-13.
[48] Bartolini, Tettamanti, and Varga, 'Critical Features of Autonomous Road Transport from the Perspective of Technological Regulation and Law', (, p. 793.
[49] Nhtsa, 'Automated Driving Systems 2.0: A Vision for Safety', *U.S. Department of Transportation.* (2017).
[50] Germany, 'Ethics Comission on Automated and Connected Driving - Report of June 2017.'.
[51] Ibid.

academics, possibly delaying the implementation or even overturning favourable laws[52]. These incidents – which will be addressed with more care further on, when analysing legal responsibility division – shine light into issues that the initial enthusiasm might obscure: that there are unexplored consequences to the adoption of a completely new way of interacting with our means of transportation.

2.4. Autonomous Warfare

When it comes to the current state of autonomous warfare, it is more difficult to ascertain at which stage different States are. Be it from the private initiative or from States themselves, the development of this technology is shrouded in secrecy[53]. Therefore, it is impossible to know how close humanity is to experiencing the first (public) deployment of LAWS.

There are, however, already some existing systems which present a high degree of automation, albeit with a more controlled and predictable scope:

> "Examples of embodied autonomous weapon systems in use today include the Israeli Harpy Loitering Weapon, an airborne weapon that identifies and destroys enemy radar emitters17 and the U.S. Aegis control system operated in conjunction with the U.S. Phalanx Close In Weapons System (CIWS), which has an operational mode that presumes human operators are incapacitated and allows it to independently identify and engage incoming anti-ship missiles and aircraft.18 Over thirty states have similar "air, rocket, and missile defense systems with human-supervised autonomous modes,"19 and innumerable increasingly autonomous weapon systems are now in development.20 Currently-deployed autonomous weapon systems act in largely predictable ways, either because they are operated in semi-autonomous modes or in relatively stable environments. For example, the South Korean SGR-A1—a stationary, armed robot used to monitor the demilitarized zone— allegedly has a fully autonomous setting, but South Korea maintains the SGR-A1 is used only in conjunction with a human operator and only in the Korean demilitarized zone." (Crootof 2018)[54]

There is no explicit International Humanitarian Law (IHL) rule specifically forbidding the use of LAWS, but according to the The Additional Protocol I of the Geneva Conventions, article 36, States are required to conduct a legal review of all new weapons, means and methods of warfare to determine whether they are compatible with international law.

---

[52] Pearl, 'Fast & Furious: The Misregulation of Driverless Cars', (, p. 21-22

[53] Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. ', (A/HRC/23/47.: UN Human Rights Council, 2013)., p.8.

[54] Crootof, 'Autonomous Weapons Systems and the Limits of Analogy'., p.7.

In 2013, the issue came to the attention of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, who produced a report on LAWS[55]. In his assessment, despite the fact that there were no known fully autonomous systems in use currently, the discussion of the matter was urgent, before the widespread use of the technology made attempts to stop it be completely fruitless[56]. His main concern was about whether LAWS would be able to comply to International Humanitarian law or International Human Rights Law, and whether the decision to kill could be left with a machine without harming the Martens Clause, with the principle of humanity[57]. On the other hand, it recognises that the substitution of human moral failings with the cold logic of robots could bring potential benefits in terms of conduct, accountability (if the systems were forced to document their actions and leave a "digital paper trail"), and, of course, with the reminder that robots do not commit rape[58].

Recognising that machines had a good potential of analysing quantitative data and exhibiting restraint, but did not fare well in calculations that involved qualitative data and contextual cues (such as the principle of proportionality, or recognising a combatant in the process of surrendering), there should be a limit not on the *technologies* that can compose LAWS, but in severely limiting the uses of such technologies to controlled situations in which the system would be able to act without much risk of violating IHL or IHRL. Even so, he recognised a gap in criminal responsibility, and proposed that the State should take on a bigger role in terms of assignment of responsibility for the use of LAWS.

In short, Heyns recognised the numerous weaknesses and potential threats posed by the automation of decision-making and especially execution of said decisions, but did not call for a complete ban on the technology – although he proposed a moratorium on the development of the technology. His desired approach would be a regulatory frame that would limit the uses of LAWS – most likely to limit the lethality itself, but leaving the field of robotics and autonomous systems free to explore other possible uses for autonomous systems (such as bomb-defusing, or even machines that could exert non-lethal force, or that were programmed to only fire after being fired at).

---

[55] Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. '.
[56] Ibid., p.7.
[57] Ibid., p.16-17.
[58] Ibid., p.10.

Meanwhile, in the sphere of non-governmental organisations, there has been extensive campaigning for banning the use of fully automated weapons (with a human out of the loop categorisation). The "Stop Killer Robots" campaign, comprised of a collective of non-governmental organisations, strives to achieve the inclusion of LAWS in the list of prohibited weapons under the Convention on Prohibitions or Restrictions on the Use of Certain Conventional weapons which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (CCW)[59]. Their continuous efforts, since 2013, have led to declarations of 26 States endorsing the ban on LAWS[60]. Other Human Rights organisations, such as Human Rights Watch[61] and Amnesty International[62] also participate in the political movement to ban LAWS.

The extent of this ban, however, is very clearly directed towards the automation of the *execution of the action* – the proposed ban is directed at fully automated weapons, whereas the presence of a human being 'on the loop' seems to appease the threshold of compliance to IHL and IHRL.

When it comes to State initiatives, even though it is not fully possible to have full knowledge about the state of development in warfare initiatives, the US Department of Defense, in their directive concerning autonomy in weapons systems, that there is no prevision of autonomous weapon systems to engage with human targets – that is, that the decision to kill can only be taken by a semi-autonomous weapon, which allows "*commanders and operators to exercise appropriate levels of human judgment in the use of force*[63]*"*.

So far, the common solution of integrating a progressively automated warfare and the demands of the regimes of Humanitarian Law and International Human Rights law seems to reside in the idea of relevant human control – meaning that, despite the rising levels of automation of all other functions, the moment of decision in engaging a target should be made by a human being.

---

[59] Noel Sharkey, 'Why Robots Should Not Be Delegated with the Decision to Kill', *Connection Science,* 29/2 (2017), 177-86.
[60] Stop Killer Robots, 'Country Views on Killer Robots.', (2018).
[61] Human Rights Watch, 'Mind the Gap'.
[62] Rasha Abdul Rahim, 'Ten Reasons Why It't Time to Get Serious About Banning 'Killer Robots'. ', *Amnesty International,* 2015.
[63] Us-Dod, 'Directive on Autonomy in Weapons Systems'., p. 7.

3. Between Automated Weapons and Automated Cars

At a first glance, the discussions surrounding driverless cars and killer robots seem to be from opposing ends of technological uses: the dream and the nightmare of how far technology can bring humankind. What could be the use, then, in tracing parallels between these two areas?

The first reason is of a practical nature: because of the differences in the way the automotive and military industries operate, the discussion surrounding automated cars operate in an environment where secrecy is much less tolerated by the public. Despite questions regarding industrial secrets when it comes to the specifics of how AI is programmed, the discussions about moral choices in car accidents, as well as re-distribution of legal responsibility are openly addressed both by the companies themselves and governments[64].

Moreover, the development of driverless cars has reached a point where tests are being performed in public roads[65], whereas the exact status of technological advancement in the military – especially in light of the controversy surrounding LAWS – is not publicised by States. The concrete issues that arise from this experimental phase, including malfunctioning, misuse and even fatal accidents, have the potential to elicit public responses and even have legal consequences that are not as of yet visible in the slower-paced development of International Law.

Many States are already experimenting legally with permissions, regulations, ethical guidelines and even incentives to the development of driverless cars technologies in their territories, whereas the Humanitarian Law and International Law response to automated warfare is slower. The way automated driving is being dealt with legally can be an interesting indicator of how specific legal questions pertaining automated warfare could be solved. But more than convenience, there is another interesting component to

[64] For example, the German Federal Government has not only appointed a panel to discuss ethical issues surrounding automated driving, but has officially adopted a set of 20 ethical principles that resulted from this discussion.

[65] Andrew J. Hawkins, 'Waymo Seeks Permission to Test Fully Driverless Cars in California', *The Verge.,* April 13, 2018.

the comparison: how the dilemmas encountered with this technology is addressed by the different legal systems.

Beyond the overarching discussions pertaining the growing use of AI, there are specific problems that have been quite prominent in the discussion related to automated driving that resonate in the discussions concerning automated warfare – and with these points of approximation, it is possible that the more open and further developed debate around driverless cars might offer valuable insights into the field of automated warfare.

3.1. Legal definition issues

In a world of laws made by and for humans, machines who have enhanced capabilities sit very uncomfortably in the current legal definitions. The general legal classification for machines – be it cars, tanks or bombs – are of tools, inanimate objects that need a human agent to give them direction and intentionality that could ensue legal responsibility[66].

The Vienna Convention on Road Traffic (1968), as was mentioned before, requires that a driver be in control of the vehicle of all times – and even with the 2014 amendment that supports automation up to level 4, there is still the requirement of a driver with the capacity to override the car's controls.

In part, this precaution can be credited to a political will towards not disturbing the current distribution of legal responsibility, by keeping a bearer of legal personhood close at hand at all times, with at least the *potential* of control and intentionality. In this way, it is possible to avoid the issue of the legal status of the AI system by direct reference to the human who has the possibility of assuming the position of a driver.

However, there is no amount of legal gymnastics that can completely obscure the higher standards that other legal fields have in determining who the subject of an action was. This works especially when determining intentionality and responsibility – mainly, the basic criminal law elements for culpability. Whereas the payment of damages in case of accident can be assigned even where there was no intentional harm, criminal responsibility demands a subjective element that would not necessarily be present when an auto-pilot function of a driverless car is engaged.

[66] Crootof, 'Autonomous Weapons Systems and the Limits of Analogy'., p.10

Of course, there is always the possibility of claiming negligent behaviour, if the driver did not comply to safety instructions about their needed level of awareness, but there might be cases in which the human being's attention span is simply not good enough to catch up and take over the driving operation in time[67].

This peculiar situation brings up several questions when it comes to judicial responsibility – and they will be discussed in more detail below. For now, it is interesting to notice, however, that the current treatment of insisting in the driver's full agency fails to capture the unique decision-making power of the machine – especially if there is a self-learning component to the AI programme, putting the predictability of its actions squarely outside the control of even the original programmers. And when it comes to the legal status of level 5 autonomous systems in the SAE scale – where there is no human driver, but only passengers – it would be ludicrous to consider a passenger, completely alienated from the driving activity, as detentive of legal responsibility.

When it comes to autonomous weapons, the issue is no less fraught with doubts and imprecisions. Rebecca Crootof (2018) explore the unique standing of autonomous weapons systems, who possess too much autonomy to fall under the definition of weapons, but lack the capability for self-determined and independent action that combatants have. Even comparisons with entities that are considered to have less capability for self-determination or independence – such as animals and child soldiers – are uncapable of capturing the unique position of weapons systems[68]. that are, at the same time, products of human artifice, but beyond the full control of their creators.

Whereas the solution suggested by Crootof (2018) is the creation of new law to regulate this new kind of hybrid weapon-agent, the current state of affairs in the international does not seem inclined to create new categories. The discussions on the CCW so far, and even the efforts of HRW and Stop Killer Robots campaigns call for a solution that would have as a result adapting automated weapons to be more easily traceable to a human controller, and not the creation of new categories. The idea of 'relevant human control', therefore, addresses the issue of legal definition by forcing technology back into the old model of how a natural human being uses a weapon – no matter how

[67] Sparrow and Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', (, p. 211.
[68] Crootof, 'Autonomous Weapons Systems and the Limits of Analogy'.

autonomous it might be in the rest of the operations – and is responsible for its deployment.

When the mere existence of this technology is already surrounded with uncertainties and challenges in its definitions, the consequences of the actions of machines guided by AI are fraught with problems, concerning both cars and weapons. When there is no human agent to be morally and legally responsible for ethical decisions, how does AI fare in situations of decision-making?

3.2. The Trolley Problem and Targeting.

The trolley problem[69] is an old moral philosophy dilemma proposed by the philosopher Philippa Foot in 1967, that is mainly used as a thought experiment to identify different ethical systems, with their advantages and shortcomings[70]. However, when it comes to the more dynamic reality of driving cars over roads, and not trains over tracks, the trolley problem seems quite divorced from reality. Human drivers in general do not feel like they are making ethical decisions of such magnitude in their everyday driving – and even when accidents or collisions occur, there is very little time for ethical considerations, and drivers end up making decisions by instinct, or just losing control of the car altogether[71].

Despite the promise of greatly reducing the number of accidents and fatalities on the road, the advent of automated driving systems cannot, realistically, completely mitigate the possibility of life-threatening situations right away[72]. The biggest difference between a human driver an AI system is that the computer will be able to substitute that split-second, instinctual decision with a pre-programmed logic encoded in the software[73].

---

[69] *"The trolley problem asks for moral justification. Is it morally permissible (...) for someone to divert a runaway trolley that would kill five people to where it will kill one different person instead, though it is not permissible to push an innocent bystander in front of the trolley, killing him, if that is the only way to save the five?"*F. M. Kamm, *The Trolley Problem Mysteries* (New York: Oxford University Press, 2015) at 2.
[70] Ibid.
[71] Noah J. Goodall, 'Away from Trolley Problems and toward Risk Management', *Applied Artificial Intelligence,* 30/8 (2016), 810-21.
[72] Sparrow and Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', (, p. 210-211
[73]Goodall, 'Away from Trolley Problems and toward Risk Management', (p. 813).

Therefore, despite the fact that the main perceived advantage of automated driving is on damage reduction, there will be situations in which the software will have to engage in a targeting decision, and have within its programme the criteria to make a decision if the situation arises.

When it comes to LAWS, the possibility of identifying and engaging targets is a much more central element instead of a mere damage-control measure, targeting would be one of the main functions of the system. But the main point of convergence of the programming of both cars and weapons is the *how*: what are the desirable parameters of the targeting decisions? How should it be programmed, and who is responsible for the killing done within these parameters?

Considering that Humanitarian Law is the most important constraint that LAWS would have to submit themselves to, the ability of following the fundamental precepts thereof is the biggest test of their legality under International Law. Therefore, LAWS would have to be able to follow the fundamental principle of distinction, as well as being able to fulfil the requirements of the principle of proportionality[74].

3.2.1. The trolley problem and proportionality

In a car crash situation where the automated system identifies that there is no safe alternative without risking harm or even death of one of the people involved, the programming will need to find a satisfactory choice between all of the parties involved – ranging from private property, animals, pedestrians, other drivers and the passengers of the vehicle involved. In any scenario, there are a multitude of interests that can be prioritised. The driver's life, who is the person who ultimately paid for the software, and is very likely to consider their own interest while choosing this or that particular system; bystanders who were not in any way involved with the collision, and should not bear the consequences of other people's mistakes; possibly vulnerable sectors of society, such as children, the elderly and people with disabilities, who might have less conditions of avoiding harm to themselves than the general population, and should therefore be protected… The list goes on, and the possibilities of balancing ethically difficult choices are almost infinite.

[74]Emily Crawford and Alison Pert, *International Humanitarian Law* (Cambridge: Cambridge University Press, 2015).

The need to balance so many different factors, weigh them against each other in advance and ultimately be accountable for them comes hand in hand with the development of technology. If the driver was not impaired in their capabilities and following the traffic rules, there can be no criminal responsibility arising from a forced decision made in a split-second[75]. There is simply no *time* for ethical considerations or choices. The choice begins when there is a programmer, maybe years before the accident itself, who made a deliberate choice towards one system of ethics that resolves the car crash situation in a particular way. But the question remains of which agent or institution should inform the decisions made by the programmer – whether the user can programme different ethical preferences into the vehicle, if State government can issue a common set of morals that should be followed by all road-users, or if the companies will have free-reign to design the decision-making system that fulfils the wishes of their market segment the best.

Laws and regulations, with their generic language, are likely to limit the parameters in which these decisions can be taken, but cannot account for every individual result, demanding more specific instructions to deal with these dilemma situations. The ethics commission assembled by the German government grappled with this issue, and arrived at the conclusion that, aside from following the principle of damage minimisation, human life should take priority over all other considerations. In this scenario, it would be preferable to damage property, or merely injure humans, rather than take a life[76].

However, even those more specific principles can also produce distortions or undesired effects in concrete situations. The ethics commission itself entertained the hypothesis of a dilemma involving killing a pedestrian instead of hitting a critical building of a nuclear power plant, which has the potential to cause far greater environmental and human damage in the long run[77].

In this sense, there seems to be way less uncertainty about which set of values has to be ultimately followed when it comes to targeting decisions in the military. The principle

[75] Patrik Lin, 'Why Ethics Matters for Autonomous Cars', in M. Maurer et al. (eds.), *Autonomous Driving: Technical, Legal and Social Aspects.* (Berlin: Springer, 2015), 69-85.
[76] Germany, 'Ethics Comission on Automated and Connected Driving - Report of June 2017.'., p. 17-18.
[77] Ibid.

of proportionality, in Humanitarian Law, is codified in article 51 (5)b. Of the Additional Protocol I of the Geneva Conventions:

> Article 51 — Protection of the civilian population
>
> 5. Among others, the following types of attacks are to be considered as indiscriminate:
>
> b) an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.

The way in which this calculation is done, however, is heavily dependent on contextual cues, and is very difficult, demanding to be judged in a case-by-case basis. The HRW report refers to proportionality as a considerably human calculation[78].

On the other hand, a machine would be able to process a larger amount of data and combine it at a faster pace than humans. Ronald Arkin (2009), a roboticist famous for his optimistic view about the employment of fully automated weapon systems, argues that machines are capable of calculating proportionality better than humans[79]. Even in face of proportionality calculation between non-interchangeable variables – the so-called "apples and oranges" calculations that are not easily programmable – could be solved by machine learning, by 'teaching' the robot through big databases of case-based reasonings, opinions of experts in the field in simulations, and other learning tools[80].

However, Sharkey (2017) argues that, despite robots potentially being able to minimise collateral damage by selecting the most appropriate weapon, or directing said weapon in the best way for that scenario, there is a different kind of proportionality that they do not have access to. Mainly, the decision of whether lethal force should be applied at a given situation in the first place[81].

The main objection of those who oppose the employment of LAWS hesitate to delegate the decision over life and death to a machine – however, there is not such a strongly worded opposition when it comes to cars, who have the *possibility* of making life and death calculations in case of accidents. They recognise the impossibility of human control – the brain has no time to make a decision, and the software is, in fact, the only one capable of making a 'targeting decision' within the very limited time frame of an unavoidable car crash.

---

[78]Human Rights Watch, 'Mind the Gap'., p.8.
[79] Ronald C. Arkin, *Governing Lethal Behaviour in Autonomous Robots* (CRC Press, 2009).
[80] Ibid., p.47.
[81] Sharkey, 'Why Robots Should Not Be Delegated with the Decision to Kill', (, p.179.

In this sense, even though the discussion about targeting and proportionality is a common thread, the two find the core of their respective dilemmas in different functions. In the case of the automotive industry, the execution of the choice – mainly, directing the car towards this or that target in emergency situations – is not what is problematized, but rather the content of the choice that the system is forced into making.

Here, the factor of speed is at the crux of the matter – when it comes to car accidents, it is the superior speed of the machine which allows it to make any kind of choice regarding targeting in situations of emergency, immediately dismissing the possibility of human intervention into the execution phase.

However, speed is also the main site of potential military advantage, and keeping a human being on the decision loop could potentially mitigate the usefulness of a semi-autonomous weapon – and, on the other hand, preserving the speed at which decisions are made could make the human decision-making a mere token gesture.

The discussion surrounding LAWS regarding proportionality is situated firmly in the moment of execution of a targeting decision, whereas the discussions of the automotive industry fixate on which criteria are used to select the target, and how the values of what constitutes the best course of action are decided, raising the question of how much – or how little – of the situation can be rectified by mitigating automation in the level of execution, in the cases where target selection can be aided, or even done entirely, by a machine.

On the subject of target selection, Roff (2014) points out how, even when there might be a human being responsible for the order of execution, that the act of selecting a target – the complex operations involving the desirability and priority of targeting – already contains enough elements of strategy, as it involves matching means to an end, and that the automation of this kind of process alone would alienate human beings not only from the moment of decision-making, but in taking part in operations that would need to change the parameters of the mission itself as a requirement for successfully making and updating targeting lists[82].

[82] Heather M. Roff, 'The Strategic Robot Problem: Lethal Autonomous Weapons in War', *Journal of Military Ethics,* 13/3 (2014), 211-27.

In this sense, while stopping killer robots might be the main means through which civil society articulates the attempt to minimise the loss of control brought by the development of AI, more attention needs to be paid by the loss of human control beyond the legally visible acts of decision-making, towards other operations, such as target selection.

3.2.2. The trolley problem and principle of distinction

The principle of distinction is outlined by art. 48 of the I Protocol of the Geneva Convention:

> In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.

In this sense, the more information the automated weapons can process, the better they could perform this task. However, modern warfare is characterised, more often than not, by difficult situations in which combatants are not clearly identified, leading to a situations where humans, even with their superior contextual knowledge, have difficulties in efficiently adhering to the principle of distinction[83]. As Heyns (2013) points out,

> The current proliferation of asymmetric warfare and non-international armed conflicts, also in urban environments, presents a significant barrier to the capabilities of LARs to distinguish civilians from otherwise lawful targets. This is especially so where complicated assessments such as "direct participation in hostilities" have to be made. Experts have noted that for counter-insurgency and unconventional warfare, in which combatants are often only identifiable through the interpretation of conduct, the inability of LARs to interpret intentions and emotions will be a significant obstacle to compliance with the rule of distinction.[84]

Some other operations - such as how to recognise a person in the process of surrendering – are even more dependant on context and social cues that might be nearly impossible to program in a robot[85]. It is almost ironic then, that the possibility of acquiring data about potential targets and making informed choices is precisely what creates a targeting problem for autonomous driving systems – in a dilemma situation, especially if there is interconnectivity between vehicles, an autonomous driving system would able to choose between targets while in possession of information about the

[83] Ibid.
[84] Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. '., p. 13.
[85] Human Rights Watch, 'Mind the Gap'., p. 8.

occupants of the vehicles that it would potentially crash into. The ethics commission acknowledged this possibility, but quite emphatically rejected the possibility of using possibly discriminatory distinctions in targeting[86].

When it comes to programmable choices and distinction of targets, there is also the additional issue of human reaction to known parameters of automated systems. In this sense, it has been theorised that the co-existence of autonomous driving systems and non-autonomous vehicles would be that human drivers could take advantage of the safety features of autonomous vehicles to purposefully engage in reckless behaviour, such as cutting in front of autonomous vehicles, knowing that their safety features will always prioritise safety[87].

3.3. Susceptibility to algorithm bias

Considering both cars and weapons can find themselves in situations where targeting decisions have to be taken, the issue of the parameters of such targeting are brought to the forefront of the discussion.

The combination between the necessity of making ethical choices, and the existence of targeting is a stark reminder that there is space for the input of value judgments. And, in the case of warfare, there is no transparency when it comes to determine which parameters are being used, especially because knowing the parameters that a lethal weapon will be programmed would render them ineffective.

When it comes to automated driving, the German ethical guidelines prohibited discrimination of any kind in targeting *persons*, but there are a host of ethical choices beyond the more drastic “trolley problem” moves that can have a biased result, if not a biased intention. One example could be in the different interactions with different types of vehicles on the road – for example, when faced with choices between crashing against an old car or a new car, even if safety is the primary concern, the result might burden one or the other social segment more heavily[88].

The same applies to the military – as much as it is easier to identify the potential for bias in the function of target selection, (and the real, very immediate consequences that

[86] Germany, 'Ethics Comission on Automated and Connected Driving - Report of June 2017.'.
[87] Sparrow and Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', (
[88] Goodall, 'Away from Trolley Problems and toward Risk Management', (

it would have), it is important to remember all the functionalities that AI has in the system. Consider the function model of information gathering, processing, choice making and action, and how it can be potentially biased to a point where it won't be even possible to trace it back to a source. After all, if there is bias in the data input phase, that bias will continue through the system since all the subsequent operations assume the input of the previous phase to be reliable.

This sort of subtlety tends to fall outside the scope of legal scrutiny – the closest determination is the non-discrimination clauses which form the cornerstone of the Human Rights regime. There are no international legal cases dealing with claims of arbitrary deaths caused by discrimination in algorithms - but depending on how subtle it is, it's not a matter of an individual killing, but a propensity to kill in a certain way, or to select targets in a certain way, that cannot be fully captured by the legal system.

3.4. Insurance and transparency

Even though car accidents and military operations have completely distinct legal fields concerned over them, they both have the assignment of responsibility as a very important component of their operations. In the case of cars, insurance claims need the assignment of blame to be very precise.

The first, most visible effect that car insurance would have with the widespread usage of AI in driverless cars is that the massive influx of data needed to operate a driverless car could also be used to ascertain the reasons and circumstances of a car crash[89]. Moreover, in cases where there is a lower-level automation system, the data collected can also be used to determine the level of risk engaged by every individual car, not in an approximate level based on social indicators, as it is now, but in the precise data that can be collected by the vehicle itself[90] – mainly, instead of calculating the insurance premium based on sex, age, place of residence of the driver and an account of where the car is parked, each insurance could be calculated based on the type of behaviour that the individual exhibits behind the wheel. When it comes to vehicles of level 5 in the SAE

[89] Maurice Schellekens, 'Self-Driving Cars and the Chilling Effect of Liability Law', *Computer Law & Security Review,* (2015), 506-17.
[90]Sparrow and Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', (

scale, it is very likely that the insurance system might shift altogether, where companies might have to be insured instead of individuals[91].

This massive amount of data produced – and the potential to process it into useful information – can be a very positive aspect for the military, which will have more precise ways to evaluate their own operation and the conduct of their individual agents. Arkin[92], while arguing for the possibility of LAWS acting as a support for human troops, mentions a possible accountability gain with the presence of a robot which could collect data on all the human counterparts[93].

However, the existence and processing of this data does not necessarily mean an increased accountability for the public in general – that would require a further step of transparency, which is not very likely to be made available spontaneously. Heyns (2013) suggests that one of the ways to increase accountability of LAWS is to force robots to keep a digital trail, so their action can be controlled. In this sense, they have the potential to be more transparent than human beings on exactly what happened[94].

3.5. Legal responsibility.

If the fact that a machine has an autonomous system, means it will be programmed to make choices, the programming of said system would have to include preferences and parameters for those choices – that is, at some point, someone will have to make the deliberate decision of *how* the machines are going to decide. Many directives are essential to the tasks of driving: the main objective of transporting goods or people from point A to point B, for example. Some of them seem pretty straightforward: it makes sense for an automated vehicle with access to traffic information to choose the less congested route. Both road traffic and warfare are legally regulated areas, so it makes sense that both autonomous vehicles and weapons should be programmed to comply with the norms of their respective fields.

The mere existence of Courts and legal disputes, however, are testimony enough to the fact that the application of laws to concrete cases is no easy, straightforward task that can be easily programmable into a machine. From Traffic Regulations to Humanitarian

91 Jeff Wargin, 'How Some Insurers Will Survive Driverless Cars', *Claims Magazine,* November 2017.
92 Arkin, *Governing Lethal Behaviour in Autonomous Robots*.
93 Ibid.
94 Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. '., p.11

Law, potentially contradictory laws are balanced or harmonised with the use of guiding principles.

While human beings are capable of exercising their own ethical preferences (and deal with their consequences, when they are in conflict with the law), the way every choice is made by a machine involves deliberate programming towards the one or the other ethical system. Therefore, there will be an outside source that will determine the ethical parameters that the machine will operate under. So this opens up the scenery for a further choice: who should have the final say in the ethical choices made by AI systems? Do autonomous systems act as an ethical proxy of the operator, therefore being programmed with individual ethics? Do developers choose this ethical approach, and if so, does this choice equate with liability?

The level of uncertainty around responsibility vary, of course, with the level of autonomy presented by the vehicle: legal responsibility, after all, is tied up to individual autonomy to choose to *act*, so as long as there is the possibility human autonomy being exercised – for example, in choosing to turn off the system and drive manually – the personal responsibility of the driver cannot be discarded.

The guidelines condensed by the commission of ethics in Germany were quite clear in assessing that a more complex, nuanced approach to legal liability would be necessary with the introduction of autonomous driving – after all, considering that the capacity of driving is the main objective of the development of a system, failure to drive properly would entail product liability.

> …not only the keepers and manufacturers of the vehicles but also the corresponding manufacturers and operators of the vehicle‘s assistance technologies have to be included in the system of liability sharing. (…) In addition, a new definition is required of the duties of care to be observed by the manufacturers, suppliers and operators of components, software and data and by developers[95].

In this sense, as long as there is the possibility of both autonomous systems and human beings driving vehicles, the legal regime applicable to accidents varies between criminal and civil responsibility of the driver and product liability of the vehicle itself[96]

95 Germany, 'Ethics Comission on Automated and Connected Driving - Report of June 2017.'., p.26.
96 Bartolini, Tettamanti, and Varga, 'Critical Features of Autonomous Road Transport from the Perspective of Technological Regulation and Law', (

However, at the current state of automation (up to level 3), the companies don't seem the least willing to share in responsibility. The most recent fatal accident involving an autonomous vehicle was from the Tesla company, in March of 2018. In their pronouncements, the company was very clear to stress the human responsibility component[97], who could have disengaged the auto-pilot setting at any moment.

In fact, as seen above, international legislation on autonomous cars makes a point of still relying on the centrality of the driver – even when the presence of the human being has only a supervisory role, they still concentrate the legal responsibility for their actions[98]. In the US, different state legislations follow in the same route, placing legal responsibility in the driver even if their control of the car is intermittent[99].

The only legislation in the world that differs in this distribution of responsibility is German law, which was amended in 2017 to legally accommodate cars with level 3 and 4 in the SAE scale, in which the driver, while still on board, may rely on the functioning of the system and will not be held liable in case of system failure when they are allowed to divert their attention elsewhere on traffic[100].

This approach of product liability, however, has the function of all but eliminating the element of criminal responsibility of the driver-turned-passenger, as well as other aspects of tort law that rely on intentionality. This might be one of the reasons why most legislation – including the international standards set by the Vienna Convention – does not completely depart from the necessity of there being a driver who could be, at the very least in theory, capable of taking over the operations of the car.

When it comes to warfare, maintaining the centrality of the operator – or rather, of a human operator – is the focal point of the 'stop killer robots' campaign, and one of the most pungent reasons for that is the enormous legal and accountability gap that is left when a human decision-maker is taken out of the picture.

---

[97] Tesla, 'What We Know About Last Week's Accident. ', (2018, 2018b), Tesla, 'An Update on Last Week's Accident'.

[98] Bradwhaw-Martin and Easton, 'Autonomous or 'Driverless' Cars and Disability: A Legal and Ethical Analysis', (, p. 8

[99] Pearl, 'Fast & Furious: The Misregulation of Driverless Cars', (, p.50.

[100] Anonymous, 'Straßenverkehrsgesetz in Der Fassung Der Bekanntmachung Vom 5. März 2003 (Bgbl. I S. 310, 919), Das Zuletzt Durch Artikel 8 Des Gesetzes Vom 17. Juli 2017 (Bgbl. I S. 2421) Geändert Worden Ist', (2017).

The solution presented in the case of cars – of treating the damages caused by automated systems in terms of product liability – are not as easily operationalized as in the case of a car accident. In a situation in which victims and families of victims of LAWS would try to pursue legal action against the manufacturer of the weapon system that employed lethal force, presenting a product liability claim in what is likely to be a foreign country is not an accessible endeavour[101].

Heyns (2013), in his report, recommended a strengthening of the role of state responsibility in these situations – however, this does not apply in the cases where non-state actors might make use of LAWS[102].

When it comes to criminal law, however, there is a very clear and insurmountable gap, once the threshold of the most basic elements of criminal law – the *mens rea* – is absent when the decision to kill comes from an entity who is not a human being. Even if the parameters of the mission were programmed by a commander – a version of the command chain in which one of the ends is not human – a commander can only be held accountable for a subordinate if they failed to take action to prevent or punish said behaviour[103].

3.6. Relevant human control

The issue of relevant human control speaks directly to the classification of autonomy by levels of human-machine interaction in decision-making: at what point, exactly, does the human delegate the task fully, and lose the possibility of having a relevant impact in the decisions being taken within each system?

From the Vienna Convention to the USDoD directives, all existing legal provisions are not ready to do away with the idea of a human being as focal point of legal responsibility – especially in the military, where the use of force can only be deployed by systems considered semi-automated[104].

However, there is a fundamental difference between the role of human control when it comes to autonomous cars and autonomous warfare. Whereas the idea of having the

[101] Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. '., p.15
[102] Ibid.
[103] Human Rights Watch, 'Mind the Gap'., p.15
[104] Us-Dod, 'Directive on Autonomy in Weapons Systems'.

human being completely out of the decision loop is politically contested quite strongly in warfare, the removal of the figure of the driver is seen as an inevitable – and highly desirable – result of the development of the technology of autonomous cars[105].

Once the dangers and undesirable effects of fully autonomous weapons have been quite exhaustively outlined, the discussion in the CCW is situated more on the spectrum of how this key concept of relevant human control can be developed in such a way that there is no risk of decisions of life and death end up being taken by non-human agents.

Heyns (2013) tentatively draws the line of this human control in the moments where value judgements have to be done[106] - but what does that mean in terms of the functionalities of the system? The main discussions about fully autonomous weapons focus heavily on the autonomy of *engagement*, without necessarily problematizing when information gathering, information processing or even the narrowing down of potential targets are handled in an autonomous way. After all, bias has the potential of being programmed all the way in the data collection process, making it less explicit and less visible, but no less present, with the added disadvantage or not being immediately perceptible by the humans who are in a supervisory role of the system.

The automotive situation in which autonomous driving functions coexist with the existence of a human driver capable of steering (with the SAE automation level up to 4) has to deal with the dynamicity of the situation between drivers and the decision loop. In this sense, the situation of driverless cars presents a concrete challenge that the warfare industry has not yet been able to answer: if all studies point out the difficulty encountered by human drivers to take back control of the car in a relevant manner, because there is too much information and things are moving too fast, what possibility would there be for a human to have effective control of a military decision if there is such a huge and complex amount of information?

And taking a step back, to the functioning of the AI system itself: there is some trade-off between *intelligence* and *intelligibility* of a system, and, from the moment that

[105] Sparrow and Howard, 'When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport.', (
[106] Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. '.

systems start acquiring the needed complexity to perform the challenging tasks that have been laid out before them – such as identifying targets or differentiating between combatants and non-combatants - the level of machine learning they will have to engage in might make them inscrutable with the amount of time that would be practical for a human being to assess the situation[107]. The ability of being easily understood by human makes it simpler, whereas the increased capacity demands more intelligence and speed than might be possible for a human mind.

In this sense, the cornerstone of legality in autonomous warfare – the existence of relevant human control – can be seen as an interesting legal solution, as it permits the human-centred legal structures to continue being used, by permitting a direct relationship between the decision being taken and the presence of a human being. There are some doubts, however, about how limited the human supervisory role can be, especially considering the nature of the systems that they are supposed to control.

.

[107] Ekbia, 'Heteronomous Human and Autonomous Agents: Towards Artificial Relational Intelligence'.

4. Conclusion

While the idea of killer robots haunts the news and is met with resistance from numerous organisations, driverless cars are hailed as a new step towards a safer, more accessible future of transportation – and every now and then, fatal accidents might remind the public that there are necessary adjustments to be made, but they do not cause such an outrage as to completely stop the advancement of the technology.

The figure of the human operator who ultimately controls the system – although the development of technology makes the workings of the system less and less intelligible – is still considered indispensable in the scenario – and the way legal systems are set up around human individuals are a solid reason to insist upon this configuration[108].

However, as was highlighted above, the more automation develops, the less actual situational awareness the driver will have, and the more artificial this responsibility becomes – and this makes it fall short of the high requirements of the criminal law system.

When it comes to autonomous warfare, the idea of "relevant human control" is an attempt to reinforce the centrality of human legal responsibility and repel responsibility vacuums that might emerge with automation, but the advantages of the development of technology might as well put the issues of legal responsibility in the same situation, that is, put it in a position in which International Criminal Law is rendered useless.

The susceptibility to bias also raises red flags: the "relevant human control" requirement only demands that automation be stopped at the "decision making level", but does not concern itself (as of yet) with the automation of all other parts of human knowledge processing – collecting data, processing data, selecting choices. With all these sides of the operation free from scrutiny, more perverse forms of illegalities – such as unlawful discrimination – might be lost under the "black box" that is the decision making system. Law is not equipped to deal with these more systematic, subtle, but not less devastating factors in decision-making. The only possible tool for countering this bleak scenario is the careful auditing and transparency in the decision-making process of cars or weapons – but this is unfortunately not very likely to happen.

[108] Schellekens, 'Self-Driving Cars and the Chilling Effect of Liability Law', (

The general public seems afraid of what the machine is going to do “autonomously” – that is, on its own – but the bigger issue is, in fact, how to truly control all the different types of input that are necessary in order for the machine to react in the way it does?

In this scenario, Humanitarian Law (and other forms of State responsibility) are seen as another possible solution for this conundrum – in the absence of human beings to hold directly responsible, it would still be possible to bring forth the State who set the mission parameters – but considering that the military sometimes buys wholesale, what is the role of companies which develop that technology? International Law – and especially Humanitarian Law – is not prepared to face the increased role of private companies in warfare, and this gap makes it so that, even if States are held responsible for the use of certain technologies, the deterring effect might not be reached if the company remains unsanctioned. This is another hole in the law that the comparison with driverless cars helps to unveil.

Heyns (2013), in his report, stresses the importance of an increase State responsibility in order to prevent the responsibility-vacuum caused by LAWS, but considering how much of the technology used in warfare is developed by private companies, heavy sanctions on States would only have an indirect impact in the development of this technology. On the other hand, a direct engagement with companies would theoretically be possible – but it is highly unlikely that victims might make claims against companies based on product liability when it comes to semi-autonomous or autonomous weapons.

Law doesn’t deal well with this complexity of multiple actors, entities and different time frames of decision-making – it presupposes autonomy in the most pure, logical sense, and this is one of the ways that AI technology challenges not only the current legal configurations, but some of the foundations of law itself. Because there is no “natural human being” to be (artificially) ascribed as a cause to their own actions, we are forced to look at the origins of each component of the decision – and be faced with interdependency, vagueness, and power struggles.

On the other hand, one could argue that all that machines are doing is injecting logic and predictability into an area in which all the imprecisions and arbitrariness was locked within the black box of the natural phenomena that is the human being. The public opinion recoils from the idea of machines making choices, but they themselves are highly dependent on a lot of ‘input’.

But in the end of the day, discussing about autonomy in machines is nothing but discussing the wider sets of conditions that we programme into them – how our humanity seeps into the code and shapes the decision. There are still human hands there – but outside of a single, biological black box of decision-making that is a human being.

Bibliographic References

Anonymous (1968), 'Vienna Convention on Road Traffic (1968) U.N.E.a.S.Council'.

Anonymous (2017), 'Straßenverkehrsgesetz in der Fassung der Bekanntmachung vom 5. März 2003 (BGBl. I S. 310, 919), das zuletzt durch Artikel 8 des Gesetzes vom 17. Juli 2017 (BGBl. I S. 2421) geändert worden ist'.

Arkin, Ronald C. (2009), *Governing Lethal Behaviour in Autonomous Robots* (CRC Press).

Bartolini, Cesare, Tettamanti, Tamás, and Varga, István (2017), 'Critical features of autonomous road transport from the perspective of technological regulation and law', *Transportation Research Procedia,* 27, 791-98.

Boulanin, Vincent (2016), 'Mapping the development of autonomy in weapon systems - a primer on autonomy', *Stockholm International Peace Research Institute Working Paper*. <https://www.sipri.org/sites/default/files/Mapping-development-autonomy-in-weapon-systems.pdf>.

Bradshaw-Martin, Heather and Easton, Catherine (2014), 'Autonomous or 'driverless' cars and disability: a legal and ethical analysis.', *European Journal of Current Legal Issues,* 20 (3).

Bradwhaw-Martin, Heather and Easton, Catherine (2014), 'Autonomous or 'driverless' cars and disability: a legal and ethical analysis', *European Journal of Current Legal Issues,* 20 (3), 1-13.

Carlson, Alyssa M. (2017), 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review,* 103 (303), 304-29.

Crawford, Emily and Pert, Alison (2015), *International Humanitarian Law* (Cambridge: Cambridge University Press).

Crootof, Rebecca (2018), 'Autonomous Weapons Systems and the Limits of Analogy', (Yale Law School).

Ekbia, Hamid R. (2015), 'Heteronomous Human and Autonomous Agents: Towards Artificial Relational Intelligence', in Jan Romportl, Eva Zackova, and Jozef Kelemen (eds.), *Beyond Artificial Intelligence - The Disappearing Human-Machine Divide* (Switzerland: Springer).

Ekelhof, Merel A. C. (2017), 'Complications of a Common Language: Why is it so hard to talk about Autonomous Weapons?', *Journal of Conflict & Security Law,* 22 (2), 311-31.

Engineering, Society of Automotive (2017), 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles', in SAE (ed.), (J3016_201609).

Germany, Federal Ministry of Transport and Digital Infrastructure of (2017), 'Ethics Comission on Automated and Connected Driving - Report of June 2017.'.

Goodall, Noah J. (2016), 'Away from Trolley Problems and Toward Risk Management', *Applied Artificial Intelligence,* 30 (8), 810-21.

Haas, Michael Carl and Fischer, Sophie-Charlotte (2017), 'The evolution of targetet killing practises: Autonomous weapons, future conflict, and the international order', *Contemporary Security Policy,* 38 (2), 281-306.

Hawkins, Andrew J. (April 13, 2018), 'Waymo seeks permission to test fully driverless cars in California', *The Verge.*

Heyns, Christof (2013), 'Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions. ', (A/HRC/23/47.: UN Human Rights Council).

Human Rights Watch, Harvard International Human Rights Clinic (2015), 'Mind the Gap - The Lack of Accountability for Killer Robots'.

Kamm, F. M. (2015), *The Trolley Problem Mysteries* (New York: Oxford University Press).

Kirkpatrick, Keith (2017), 'It's Not the Alrorithm, It's the Data - in risk assessment and predictive policing, biased data can yield biased results.', *Communications of the Association for Computing Machinery,* 60 (2), 21-23.

Lin, Patrik (2015), 'Why ethics matters for autonomous cars', in M. Maurer, et al. (eds.), *Autonomous Driving: Technical, Legal and Social aspects.* (Berlin: Springer), 69-85.

NHTSA (2017), 'Automated Driving Systems 2.0: a vision for safety', *U.S. Department of Transportation.*

Parasuraman, Raja, Thomas B. Sheridan, and Wickens, Christopher D. (2000), 'A Model for Types and Levels of Human Interaction with Automation', *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans,* 30 (3), 286-97.

Pearl, Tract Hresco (2017), 'Fast & Furious: The Misregulation of Driverless Cars', *NYU Ann. Surv. Am. L.,* 73 (19).

Rahim, Rasha Abdul (2015), 'Ten reasons why it't time to get serious about banning 'Killer Robots'. ', *Amnesty International*.

Richards, Dale and Stedmon, Alex (2016), 'To delegate or not to delegate: A review of control frameworks for autonomous cars', *Applied Ergonomics,* 53, 383-88.

Robots, Stop Killer (2018), 'Country Views on Killer Robots.'.

Roff, Heather M. (2014), 'The Strategic Robot Problem: Lethal Autonomous Weapons in War', *Journal of Military Ethics,* 13 (3), 211-27.

Schellekens, Maurice (2015), 'Self-driving cars and the chilling effect of liability law', *Computer Law & Security Review*, 506-17.

Sharkey, Noel (2017), 'Why robots should not be delegated with the decision to kill', *Connection Science,* 29 (2), 177-86.

Sparrow, Robert and Howard, Mark (2017), 'When human beings are like drunk robots: Driverless vehicles, ethics, and the future of transport.', *Transportation Research Part C,* 80, 206-15.

Tesla (2018a), 'An Update on Last Week's Accident'.

--- (2018b), 'What we know about last week's accident. ', (2018).

UNECE (2016), 'UNECE paves the way for automated driving by updating UN international convention'.

US-DoD (2012), 'Directive on Autonomy in Weapons Systems', in US Department of Defense (ed.).

Wargin, Jeff (November 2017), 'How Some Insurers will Survive Driverless Cars', *Claims Magazine*.

Zackova, Eva (2015), 'Intelligence Explosion Quest for Humankind', in Jan Romportl, Eva Zackova, and Jozef Kelemen (eds.), *Beyond Artificial Intelligence - The Disappearing Human-Machine Divide* (New York: Springer), 31-44.