UNIVERZITA KARLOVA

Právnická fakulta

Andreas Nanos LL.M.

Srovnávací Analýza Trestní Odpovědnosti Autonomního Řízení a Silné Umělé Inteligence

Disertační práce

Školitel autora disertační práce:
prof. JUDr. Bc. Tomáš Gřivna, Ph.D.

Theoretical Legal Science –
Law and Legal Theory in European Context

Datum vypracování práce (uzavření rukopisu): 2023-08-16

CHARLES UNIVERSITY

Faculty of Law

Andreas Nanos LL.M.

Comparative Analysis of Criminal Liability of Autonomous Driving and Strong Artificial Intelligence

Dissertation

Supervisor: prof. JUDr. Bc. Tomáš Gřivna, Ph.D.

Programme: Theoretical Legal Science –
Law and Legal Theory in European Context

Date of dissertation (closure of manuscript): 2023-08-16

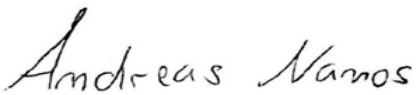
Statutory Declaration

I declare that I wrote the submitted dissertation independently and that all the sources were duly stated and that the dissertation has not been used to attain another or the same degree.

Furthermore, I declare that the actual text of the dissertation, including footnotes, has 324.166 characters including spaces.

Prague, 2023-08-21

Place, Date



Andreas Nanos LL.M.

Table of Content

Introduction

'It is not my fault it is the algorithm!' This excuse is commonly heard whenever a digital application malfunctions in everyday use. While it may be true that a faulty algorithm meant artificial intelligence (AI) caused an unwanted event that injured the legally protected interests of human beings, the question of liability remains unanswered. In such cases, finding the liable person is essential and, when it comes to AI, the question reaches an unexpected depth, since any AI-applications involve a multitude of different parties – all playing their own distinct and equally important role.

Even though we are still decades away from AI which is as intelligent as human beings (strong AI), AI in general already has a significant influence on our society. However, in saying this, much of this influence is not currently noticeable, since AI systems tend to be in the background, working behind applications and machines that provide us with functions which would not be possible without AI.

In terms of benefits for the future of mankind, the promises of AI are extraordinarily high. Yet, at the same time, the rapid pace of development in this rather new technology is matched by increasing concern and anxiety from the general public. Without doubt, many people are afraid of AI, be it as a result of dystopian literature and movies such as 'The Terminator', in which intelligent robots try to eliminate the entire human population, or because they fear the social changes AI might bring, such as degeneration in social interaction or the loss of employment. This fear is certainly not unfounded: it is said that AI-systems may potentially perform any task usually undertaken by human beings, even if the specific tasks or jobs require years of professional training. This implies an immense potential for changes of our society and every-day life, since it means that every single existing job may be done by a robot in future.

Adaptive intelligent agents open up new dimensions of technological progress. Simultaneously, the legal evaluation of their behaviour results in highly complex questions, especially the question of liability. Although it should be underlined that no final solution to the problem has been found to date, scholars

in civil law have researched the problems of liability stemming from AI quite extensively. Criminal law, however, is still far behind. Where legal interests relevant to criminal law are violated, for example resulting in the death of a person or the destruction of property, questions concerning criminal liability arise immediately, including: who acted wrongfully and who is criminally liable? Is it the human being who put the machine into operation or the intelligent agent itself? Or, alternatively, should nobody be held criminally liable for the agent's deeds because society recognises the use and development of AI and would rather bear the consequences of its partial unpredictability?[1]

German criminal law, for example, is anthropocentric. The eventuation of the misconduct is always attributed to a human being if he caused the offence by his deeds. Further, a human being is criminally liable for fulfilling the factual requirements of intention or negligence, while being able to recognise his wrongdoing and his possibility to reasonably prevent it.[2]

However, this thesis will further consider AI, also known as intelligent agents, of being capable of learning and, to some extent, acting autonomously. Therefore, allocating liability is one major concern in this regard. A person who is responsible for any damage shall be liable and obliged to compensate for it. In cases, where only one person is responsible, allocating responsibility is often relatively easy. However, finding the responsible person becomes much harder if there are several parties involved, where the different work steps and tasks are distributed, or in value chains where different services are provided. The advent of any meaningful new technology usually leads to questioning the current rules of liability are frequently questioned, and they are often thereafter defined accordingly. A relevant example is the introduction of intelligent agents, which are able to process information from their environment, and act accordingly, without the intervention of human beings. However, the use of such agents and,

[1] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, p. 565.
[2] Ibid.

for example, autonomous driving on a large-scale has sparked legal discussion concerning liability for such robots.[3]

Allocation of responsibility is a challenge within the use of AI, hence there are a wide range of different liability models that are potentially applicable in the variety of technologies equipped with AI. AI-systems may involve a multitude of individuals, including, for example, the programmer, producers, producers of individual components in the machine, and users.

The following research considers a set of possible accidents, and other cases which show the potential threats and risks of AI to the interests of individuals and collective interests and values. In doing this, the focus is on an assessment of the legal basis, especially in terms of criminal liability. Since machines are likely to be involved in substantial and systematic wrongdoing, including the violations of human rights[4] and may even commit such themselves, criminal liability is an essential legal question which needs to be examined. As such questions have already begun to play a role in some courts, the current development of AI suggests that the importance of a stable legal basis will improve exponentially in the future.

AI occurs in manifold different peculiarities. Just as natural intelligence, many types exist. However, the different types do not form part of this research, which rather narrows the differentiation to being between strong and weak AI. Strong AI is defined as an AI that is at least as intelligent as humans. Under the definition of weak AI, systems are classified as being designed solely for the purpose of fulfilling one or a limited number of specific tasks and therefore are not considered "intelligent" in the proper sense.[5] As the research will show, this distinction is important since it has major impact on (criminal) liability.

This study represents a comparative analysis between strong and weak AI in terms of criminal liability. Since there are not many cases in which AI is

---

[3] Gless, S./Seelmann, K. *Intelligente Agenten und das Recht – Verantwortungszuschreibung in Antike und Moderne.* In. *Intelligente Agenten und das Recht, Robotik und Rec*ht, 1st edition, vol. 9, Baden-Baden, Nomos, 2016, p. 11.

[4] Yeung, K. *Responsibility and AI: Council of Europe Study,* DGI(2019)05, Council of Europe, 2019, p. 44.

[5] For more detail on the difference between 'strong' and 'weak' AI see Part. 2, and Part 3.

appropriate to criminal law, existing and especially established criminal law doctrines are analysed and, where appropriate, applied to AI. Strong AI has never been subject to a criminal or any other court in history since it does not yet exist. Therefore, this research especially relies on the application of traditional criminal law doctrines. As regards strong AI, the research is mostly based on fictional scenarios and technologies which might exist in the future. Of course, it could be said that scholars should focus on the legal basis of technologies that are already existent, or at least on the verge of existence. However, in light of the immense impact and the possibilities of strong AI, this thesis takes the position that it is advisable for scholars (of all subjects) and law makers to approach this topic as soon as possible. This is the only way to prepare society for the introduction of the new, ground-breaking, and society-changing technologies.

However, not only legal aspects are significant for this research. In particular, a brief examination of moral aspects should be undertaken concerning the question of how artificial intelligence should react in certain situations (e.g., dilemmas).[6]

Unfortunately, one question, potentially the most difficult question surrounding this topic, goes beyond the scope of this thesis. That is, if an intelligent system itself (as an autonomous agent) can be subject to criminal law, how should such a machine be punished in a way that the system 'feels' it as such? This question goes far beyond the competence of legal studies. A machine which cannot feel and has no perception of time, probably cannot be punished.

Entities which have a sense of morals, are able and free to make choices concerning their decisions and actions. However, their actions and decisions can be wrong and may even cause harm to other people's health, property or other legal interests.[7] Can an intelligent agent only be considered responsible if it is

[6] See Chapter 9, Section 9.6.
[7] Yeung, K. *Responsibility and AI: Council of Europe Study,* DGI(2019)05, Council of Europe, 2019, p. 47.

aware of its environment and the influence of its actions, and if it can decide freely from a variety of possible alternative actions based on its environment?[8]

This thesis does not directly focus on any specific legislation. It does, however, take national laws (mostly German and Swiss) as examples. The research purely analyses how criminal laws already deal or might deal with AI in future. The research shall, however, especially in terms of strong AI, provide ideas of how criminal law might be applicable. Therefore, as concerns strong AI, factors such as cognition and sensory capabilities are added to the research to make it a more fitting subject to criminal law. The research shall further show which legal prerequisites are to be fulfilled so as to make AI criminally liable.

By developing autonomous robots with the ability to learn, we are building machines that take on responsibilities even at the stage of decision-making. This can be characterised as technological reaction to the over-complexity of modern society, in which one not only has to make numerous decisions everyday but also knows that many decisions carry the potential to harm others. Because of this we are building machines not just to decide how to best find our way in traffic or to get our car into a parking spot, and not just to remind us about when we should take our medicine or buy food once the fridge is empty – we are building machines to decide about life and death of other human beings. Take persons having to decide if the medial doctor shall turn off life-sustaining measures of their closest relatives. The chances of recovery are low, but not zero. Neutrally weighing up all options, purely based on the given facts, without the influence of emotions is barely possible. Without a doubt, decisions about life and death of other human beings are very hard to make for many people. Making such decisions can be emotionally overwhelming. It is suspected that, behind this reason, lies the need to hand over such difficult decisions to autonomous robots, which are supposed to be able to make such decisions without the problems of being emotional or overwhelmed by the given situation.

It is suspected that, behind this development, lies the need to hand over such decisions because human beings feel overwhelmed by the responsibility for

[8] Matthias, A. *The responsibility gap: Ascribing responsibility for the actions of learning automata*. In. *Ethics and Information Technology.* 6(3), 2004, p. 175.

them. There is the hope that machines might make fewer mistakes by having more information and reacting faster than human beings. But this development leads to two questions: who is the responding entity? Can the machine respond in a way that is necessary for social and legal responsibility?[9]

As this thesis will demonstrate, AI and artificially intelligent technology carry multiple threats and risks to individuals and society-related interests and values. Further, such technologies may facilitate the commission of wrongdoing, including the violation of human rights (in the sense of criminal law as well). In summary: these technologies may threaten health, collective moral and social foundations of democratic societies. In order to compare the criminal liability of strong and weak AI, the following research seeks to clarify who needs to be held criminally responsible if AI-systems harm, or violate the rights of individuals, groups or society as a whole. An additional inquiry in this study pertains to whether the existing criminal law doctrines are adequate to address offenses involving AI, or if a novel specialized criminal law is indispensable.

1. Terminology

1.1. Intelligence

The paper "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence"[10] is commonly recognised as the terminological emergence of AI.[11] Intelligence in its most general form, according to Legg and Hutter, can be defined as follows: "intelligence measures an agent's ability to achieve goals in a wide range of environments."[12] In this manner, the term 'intelligence' is positioned as a characteristic. That is, the acting entity has 'intelligence' (or not).[13]

---

[9] Beck, S. *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 140.
[10] By J. McCarthy, Dartmouth College M. L. Minsky, Harvard University N. Rochester, I.B.M. Corporation C.E. Shannon, Bell Telephone Laboratories, from August 31, 1955.
[11] Herberger, M. *"Künstliche Intelligenz" und Recht.* In. *Neue Juristische Wochenschrift.* 39/2018, p. 2826.
[12] Legg, S./Hutter, M. *Universal Intelligence: A Definition of Machine Intelligence*. In. *Minds & Machines*, 17(4), 2007, p. 12.
[13] Herberger, M. *"Künstliche Intelligenz" und Recht.* In. *Neue Juristische Wochenschrift.* 39/2018, p. 2826.

Besides this, another approach understands 'intelligence' as a relation term. This definition aims to attribute intelligence based on social actions. It defines intelligence as such when intelligence is recognisable to human beings within a machine-human interaction. Thus, it is intelligent when 'we recognise' something as being intelligent. Here, there is no need to attribute a technical system, (used as assistant for problem solving) intelligence per se. Its intelligence manifests itself through its interaction with human beings.[14]

1.2. Artificial

The basic meaning of the term artificial is defined as "humanly contrived often on a natural model: man-made."[15] This definition stresses the fact that it is about a man-made entity. Etymologically, the term 'artificial' originates from the Latin word 'ars'. Due to this, 'artificial' not only constitutes a man-made entity, but also a 'workmanlike' made entity. To be able to do this, the designer as 'artifex' must possess the required skill.[16]

1.3. AI and Robots

A uniform definition of the term robot does not exist.[17] AI is, in a broader sense, a computer-based system, able to solve complex problems, or act adequately to achieve its goals in possibly any circumstances it encounters. Some experts suggest varying taxonomies of problems regarding AI on one hand, and the specific solution on the other. Russel and Norvig applied the following taxonomy: 1. AI that thinks in a manner highly similar to human beings 2. AI which behaves in a similar manner to human beings. 3. Logically

---

[14] Görz, G./Nebel, B. *Künstliche Intelligenz.* 1st edition, Frankfurt am Main, S. Fischer, 2015, p. 11.

[15] Merriam-Webster, entry 'artificial', https://www.merriam-webster.com/dictionary/artificial.

[16] Herberger, M. *"Künstliche Intelligenz" und Recht.* In. *Neue Juristische Wochenschrift.* 39/2018, p. 2827.

[17] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 596; Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 02/2017, p. 173; Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 38; Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 6.

thinking systems 4. AI that is able to behave in rational ways.[18] Further, venture capitalist Frank Chen separated the problem space of AI into five different groups: logical reasoning knowledge representation, planning and navigation, natural language processing, and perception.[19] What AI actually *is*, however, is difficult to define. This difficulty stems from the diversity of AI problems and solutions, and the foundation of AI in human evaluation concerning the performance and accuracy of the applied algorithms. One example is the analysis of large amounts of data which was based on techniques, originally developed by AI researchers. These techniques are now considered as 'Big Data' algorithm systems. Challenges in defining what represents AI origins in the vastness of the issues and solutions required to be solved by AI, as well as the underlying performance of algorithm fuelling the development of AI.[20] Still, the boundaries may happen to be uncertain, just as they may tend to shift. What is important, however, is the fact that one of the core objectives of AI research and applications has been the automation or the replication of intelligent behaviour.[21] The designation was first introduced by the Czech author Karel Capek in the year 1920, relating to artificial human slaves, who ultimately wipe out humankind.[22] Some scholars define robots as sensorimotor machines in order to extend the human capacity for action.[23] These machines stand out from conventional machines through their higher complexity, superior autonomy and more complex software.[24] These machines are composed of mechatronic components, sensors and computer-based control functions.[25] Bekey defines

[18] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 6. (Citing Russel, S./Norvig, P. *Artificial Intelligence: A Modern Approach.* 3rd edition, Hoboken, Prentice Hall, 2009).
[19] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 7.
[20] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 38.
[21] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 7.
[22] Christaller, T. et al., *Robotik und menschliches Handeln.* In. *Robotik. Wissenschaftsethik und Technikfolgenbeurteilung.* vol. 14, Berlin, Heidelberg, Springer, 2001, p. 18.
[23] Ibid., p. 5.
[24] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 173.
[25] Christaller, T. et al., *Robotik und menschliches Handeln.* In. *Robotik. Wissenschaftsethik und Technikfolgenbeurteilung.* vol. 14, Berlin, Heidelberg, Springer, 2001, p. 5.

robots "as a machine, situated in the world, that senses, thinks, and acts."[26] This means that robots require sensors and the computation-force to mimic certain parts of cognition. Sensors are necessary to receive information from the environment. Robots need processors to gain certain cognitive abilities, and actuators to enable the robot to act in its environment.[27] However, literature does not provide consistent or uniform definition of robots. According to the official guideline of the German engineer's association,[28] robots are universally applicable moving automates with several axes, whose movements and sequence of movements are freely programmable and, where necessary, are controlled by sensors.[29] All of the definitions share the common feature that they refer to a (usually sensorimotor) machine, that serves to extend human capabilities and stands out from conventional machines due to its increased complexity, owing to its more complex software.[30]

Regardless of the existence of a sensorimotor machine capable of performing certain movements and thus imitating human actions, legal responsibility is no longer based solely on this ability to act, but rather on the system's capacity to make decisions, which are brought about by the system's "intelligence" and exhibit a certain degree of autonomy. Legally relevant in terms of liability (also criminal liability) is especially the capacity of decision-making initiated by the intelligence of the robot's software.[31] For this reason, various scholars include not only intelligent moving machines into their definition of robots, but also those systems capable of autonomous decision-making, independent from any abilities of sensory movements.[32]

---

[26] Bekey, G. A. *Autonomous Robots – From Biological Inspiration to Implementation and Control,* MA, London, The MIT Press Cambridge, 2005, p. 2.
[27] Ibid., p. 18.
[28] For more information see: https://www.vdi.de/richtlinien
[29] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 173.
[30] Ibid.
[31] Ibid.
[32] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, p. 562 ff.; see e.g. Müller, M. F. *Roboter und Recht - Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 595 ff.; Wildhaber, I. *Die Roboter kommen – Konsequenzen für Arbeit und Arbeitsrecht.* In. *Zeitschrift für Schweizerisches Recht*, 135(4), 2016, p. 315 ff.

1.4. Autonomy

The concept of individual autonomy is seen as one of the fundamental concepts used to justify criminal laws. According to this, individuals are responsible for their respective deeds and shall be treated as such. Individual autonomy has two elements: the factual and normative.[33] Both elements require at least a brief examination to clarify the role of autonomy in the criminal liability of AI.

Both elements require at least a brief examination to clarify the role of autonomy in the criminal liability of AI. The factual element constitutes the individual's general capacity and free will to make significant choices. The truth of this statement has never been proven scientifically, however. The 'free will' argument, which dominated for centuries faces increasing opposition by statements of determinists, which hold that people's actions are dependent on several factors and are therefore determined, since these factors cannot be controlled. It describes human behaviour as determined by a variety of causes not controlled by the individual. The majority of philosophers tend to adopt a middle ground where they acknowledge the basic premise that behavior cannot be entirely predetermined, making it unjust and unsuitable to assign blame in most cases. They recognise that human decision-making is a complex interplay of internal and external factors, and the notion of complete freedom of choice might be an oversimplification. However, they also recognize that there are situations where behaviour can be heavily influenced, such as when one is threatened by another, leading to a displacement of the usual presumption of free will.[34]

Translating these discussions of individual autonomy to AI presents unique challenges. Unlike humans, AI lacks subjective consciousness and emotions, leading some to argue that the notion of autonomy cannot apply to these systems at all. However, it must be kept in mind that most everyday actions are performed with the belief that the individual is indeed responsible for these. Thus, where

[33] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th edition, Oxford, Oxford University Press, 2013, p. 23.
[34] Ibid.

the proof of determinism is absent, it would be wrong to abandon the assumption of free will, which influences many of our social practices. The capacities which the law assumes, though, are probably absent in children or mentally disabled people. Scholars relate to these capacities in a sense of 'preconditions of criminal liability'. In consequence, according to the assumption of the law, only sane adults may be held liable for their conduct, and matters under their control, providing there are no present exceptions, such as duress, mistakes etc.[35]

The normative element is an equally important component of the principle of autonomy. It states that individuals are to be considered as agents, capable of choosing what they do and what they omit to do.[36] These abilities are fundamental as far as the idea of individuals being capable of choosing the nature of their conduct is concerned.[37] According to Joseph Raz, three main features are used to characterize the autonomy-based doctrine of freedom. The first feature is promotion and protection of positive freedom. This is understood as the ability for autonomy, which is composed of a variety of an adequate range of options available, and the mental abilities needed to be able to live an autonomous life. The second feature is the state's duty to promote freedom and to prevent the denial of it. This goal is to be reached through the creation of the conditions for autonomy. Third, individuals may not follow objectives in a sense which violate another individual's autonomy unless there are circumstances which justify such actions in order to protect or promote the autonomy of those individuals or others.[38]

According to H. L. A. Hart's principle of individual autonomy, which is highly recognised in this regard, individuals shall be held criminally liable, if they have the capacity and reasonable opportunity to do otherwise.[39] Feinberg argues that the most basic autonomy-right is the right to decide how an individual lives his or her life, and especially the right to decide how to make critical

[35] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th edition, Oxford, Oxford University Press, 2013, p. 4.
[36] Ibid., p. 3.
[37] Ibid., p. 4.
[38] Raz, J. *The Morality of Freedom.* Oxford: Oxford University Press, 1988, p. 104.
[39] See Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd Ed., New York: Oxford University Press, 2008, Chapter 6.

decisions, such as courses of study, what skills to develop, what career to pursue etc.[40]

Autonomy in robotics can be understood to mean as little as the ability to perform tasks without continuous human input or control. Proponents of AI autonomy argue that although AI may not possess consciousness, they can still exhibit a form of decision-making that is based on their programming and interactions with data, leading to outcomes that may not be fully predictable or controllable by their human creators. This perspective emphasizes that AI systems can act independently within the bounds of their programming and learning capabilities, even if they lack subjective experiences. Autonomy, however, is defined as the ability to make (limited) decisions about what behaviours to execute based on perceptions and internal states, rather than following a pre-determined action sequence based on pre-programmed commands. Autonomy concerns the attribution of meaningful control. Meaningful control is 'based' on power and insight. Without these prerequisites, meaningful control over carrying out operations and actions is impossible.[41]

1.5. Difference between Autonomy and Automation

Often, AI is applied in systems, able to control physical machines or make actions online. Issues of autonomy, automation and human-machine teaming may arise when AI is allowed to come into contact with the everyday world.[42] The term autonomy describes the ability of systems to operate and adapt to varying environments. A good example here relates to self-driving cars, where AI algorithms enable vehicles to navigate through traffic, make decisions, and respond to changing road conditions without direct human intervention. Even though majority of the literature is heavily focused on autonomous vehicles, the concept of autonomy and AI is much broader and contains an enormous variety of possible examples, such as automated trading on stock markets, where AI

[40] Feinberg, J. *Harm to Self.* In. *The Moral Limits of the Criminal Law.* vol. 3, Oxford: Oxford University Press, 1986, p. 54.
[41] See Scheutz, M/Crowell, C. R. *The Burden of Embodied Autonomy: Some Reflections on the Social and Ethical Implications of Autonomous Robots.* In. *Workshop on Roboethics at the International Conference on Robotics and Automation.* 2007, p.1.
[42] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 10.

systems autonomously analyse market trends and execute trades based on predefined strategies. Further, it includes systems able to diagnose and repair faults on their own as well. A suitable example, therefore, would be the identification and fixing of vulnerabilities in security-systems[43] such as firewalls etc. AI-driven security systems can continuously monitor network traffic, detect potential threats, and autonomously implement countermeasures to protect against cyberattacks.

Automation, on the other hand, describes machines doing work that had previously been performed by a person. Several definitions exist here, such as the substitution for human labour, or systems functioning with low or without human intervention. The term is applied to both physical tasks, where machines take over manual operations, as well as mental or cognitive work, which AI-powered systems handle with increasing proficiency. The growing capabilities of AI have enabled automation to extend beyond repetitive physical tasks to complex decision-making processes, data analysis, and even creative tasks, making it a transformative force in various industries.

However, the impact of automation on employment is nothing new. Indeed, since the industrial revolution, the effects of automation have, at the very least, had a significant social and economic impact. While automation has historically led to the displacement of certain job roles, it has also given rise to new opportunities and industries, creating a net balance of job gains and losses over time. With each wave of automation, certain tasks became more efficient, leading to increased productivity and economic growth.

It is generally accepted that AI will automate or replace some jobs, as it already has. For instance, in manufacturing, robots have taken over repetitive assembly line tasks, while in customer service, chatbots have automated responses to common queries. The much wider disputed question is whether AI is just a further step in automation or whether it will affect the economy in different ways to the waves of automation in the past. [44] Unlike previous waves

[43] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 10.
[44] Ibid.

of automation, which primarily targeted routine and manual tasks, AI's capabilities extend into areas that involve complex decision-making, problem-solving, and pattern recognition.

The concern is that AI-driven automation may impact a broader range of jobs, including those that were traditionally considered safe from automation, such as certain white-collar professions like legal research, data analysis, and medical diagnostics. This broader reach of AI automation could potentially result in more significant disruptions in the job market than previous technological advancements.

1.6. Algorithms

The subject of attribution of "artificial intelligence" is usually the algorithms. Sometimes this leads to the impression that these algorithms are independent entities equipped with AI. They are frequently attributed to autonomy, which is indicated by metaphors such as ‘autonomous driving’.[45] At first consideration, the above-mentioned description displaying the mechanism of the algorithms used for decision-making may be correct. Algorithms can perform the relevant actions because human creativity established them and gave them their specific role. Thus, these algorithms are the products of human minds.

However, this does not contradict the fact that the human originator is possibly not always able to predict the outcome calculated by his own created algorithm. This lack of perfect predictability arises due to various factors, including the vast amount of data that AI algorithms process and the intricacies of their self-learning capabilities. This does not change the fact that the creator implemented the procedure, which leads to a surprising outcome. The dynamic nature of AI algorithms can result in novel solutions and insights that even their human creators may not have envisioned during the development stage.

When talking about algorithms making decisions, it must always be kept in mind that they only make decisions through delegated decision-making

[45] Herberger, M. *“Künstliche Intelligenz” und Recht.* In. *Neue Juristische Wochenschrift.* 39/2018, p. 2828.

capacities, and solely within the specific scope of autonomy they are granted. [46] The decisions made by AI systems are based on predefined rules, patterns, and objectives set by humans. However, this might be a matter of fact regarding AI that is not equipped with the ability to learn, to program, or even develop algorithms by itself. In such cases, the AI's decision-making capabilities remain limited to the programmed rules and data it was initially provided. In this case, the above-mentioned statements might be obsolete, as the AI's actions are solely determined by its programming and lack the capacity to autonomously adapt or evolve.

1.7. Accountability

Accountability means that an acting entity is liable for its actions (subject to giving an account: answerable) or that the action of an entity is explainable (capable of being explained: explainable).[47]

According to Kroll, "a process or entity is accountable for a decision if that process is consistent with the legal, political, and social norms that define its context, and this fact is publicly evident, either because such consistency can be readily determined using public information or because an entity empowered and trusted to define such consistency (e.g., a court, a legislature, or other designated authority) can be shown to have the necessary information to determine this consistency."[48]

1.8. Autonomous Agents

One of the most oft-cited definitions of agents essentially describes the definition of autonomy:[49] "an agent is a […] system situated in some environment, and that is capable of autonomous action in this environment."[50] A

---

[46] Herberger, M. *"Künstliche Intelligenz" und Recht.* In. *Neue Juristische Wochenschrift.* 39/2018, p. 2828.
[47] Kroll, J. *Accountable Algorithms*. Princeton, Princeton University Doctoral Dissertations, 2015, p. 56 ff.
[48] Ibid., p. 57; see also Kroll, J. et al. *Accountable Algorithms.* In. *University of Pennsylvania Law Review.* 165(3), 2017, p. 633.
[49] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 569.
[50] Wooldridge, M. *An Introduction to MultiAgent Systems.* 2nd edition, Hoboken, John Wiley & Sons, 2009, p. 21.

certain grade of self-control makes agents uncontrollable to some extent by outside agents. Still, the basis of autonomy or autonomous behaviour is motivational evaluation, containing the agent's assignment of utility between a number of different behavioural alternatives. After signing the utilities, the agent can make a decision.[51] Based on this idea, some scholars segment the concept of autonomy into two elements: a) minimum capability of self-government; b) minimum capability to take a value-oriented decision.[52] The first capability refers to the ability to act without the control from an external entity. The second is connected to the possibility of an agent to determine its actions according to a sort of axiological order.[53] Following this logic, another distinction has been developed: a weak definition of autonomy is proposed to the extent to which a machine is able to perform its own processes and operations without external control, while a strong definition of autonomy is proposed to be the scale in which the machine is able to sense the environment, plan, and act upon that environment, intending to reach a specific goal with little or no external control.[54]

Generally, there is a tendency when defining artificial, autonomous agents to involve both software and embodied hardware. This description is very imprecise, however, meaning that, at first, a focus in artificial life is necessary to identify the class, as the authors Ianni and Monterossi refer to a 'sub class' of AI.[55] At this point it needs to be mentioned that, dividing autonomous agents into a 'sub class' or 'sub category' is the wrong approach. For a certain level of intelligence, it might be correct to put them into different categories, however as

[51] McFarland, D/Bösser, T. *Intelligent Behaviour in Animals and Robots,* MIT Press 1993, p. 213.
[52] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 570. The authors refer to: Tzafestas, S. G. *Roboethics. A Navigating Overview.* Berlin, Springer. 2016, p. 196.
[53] Ibid.
[54] Beer, J. M. et al, *Toward a framework for levels of robot autonomy in human-robot interaction. In. Journal of Human-Robot Interaction,* 3(2), 2014, 74-99, citing Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 570.
[55] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 571.

soon as autonomous agents reach or even outpace human intelligence, the use of the word 'sub' categories is misplaced.

The term 'Artificial life' refers to the creation of existence by human intelligence, unlike natural evolution. In this context the definition of artificial life refers to systems, which are able to implement lifelike behaviour through their combination of soft-and hardware.[56] This means that systems equipped with such a degree of AI may achieve a full, autonomous existence. From the very moment at which artificial agents gain the ability of self-determination and develop relationships with other subjects of artificial and natural life, they are likely to be defined as artificial agents.[57] Some scholars have created an adequate definition of autonomous artificial agents: "an 'autonomous' artificial agent possesses a relatively higher degree of one or more of the following: the ability to operate without the direct intervention of humans or other agents, and to exert non-supervised control over its own actions and internal states; the social ability or capacity to interact with other artificial agents or with human beings; the proactive ability to initiate goal-directed behaviour; the reactive ability to perceive an environment and respond to changes within it; the ability to adjust to the habits, working methods, and preferences of users, other agents, or humans; the ability to move around a virtual or physical environment; and representativeness, or the attribute of being a representative of, or an intermediary for, another agent or person."[58]

Depending on their intellectual, cognitive and relational abilities, artificial autonomous agents may be identified as such when they have a minimum degree of AI. Thus, it makes sense to differentiate between weak and strong AI.[59] The difference between these two types is that weak AI only acts as if it were

[56] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 571.
[57] Ibid.
[58] Chopra, S./White, L. F. *A Legal Theory for Autonomous Artificial Agents.* University of Michigan Press, 2011, p. 10.
[59] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 571.

intelligent (weak AI hypothesis), whereas strong AI can actually think and not only *simulate* thinking (strong AI hypothesis).[60]

1.9. Deep Learning

In the recent years, some of the most significant advancements in the field of machine learning have been made in the subfield of deep learning, also known as network learning. It is inspired by the structures of the human brain. Deep learning structures use a set of units that 'imitate' neurons. In order to produce an output value, each unit combines a set of input values, which are then passed to other neurons downstream. This architecture allows deep learning models to perform hierarchical feature learning, gradually capturing more abstract and complex representations as data flows through the network.

To show a practical example of an image recognition application, raw data of one specific image might be combined by a first layer of units to recognise simple patterns. The results of the first layer might be combined by the second layer, to recognize patterns-of-patterns. Finally, the results of the first and second layer may be combined by a third layer etc.[61] This process of hierarchical feature extraction continues through multiple layers, allowing the network to identify increasingly intricate details in the image. Each layer of the network contributes to a deeper and more comprehensive understanding of the image, leading to accurate and sophisticated recognition. Typically, deep learning, or neural networks, use a variety of layers combined with a high number of units in each layer to enable these networks to recognise extremely complex and precise patterns in data.[62] The ability to handle vast amounts of data and identify intricate patterns has propelled deep learning to achieve ground-breaking results in various domains, such as computer vision, natural language processing, speech recognition, and more.

As the field of deep learning continues to evolve, researchers and practitioners strive to enhance the efficiency and scalability of these models

[60] Russel, S./Norvig, P. *Artificial Intelligence: A Modern Approach.* 3rd edition, Hoboken, Prentice Hall, 2009, p. 1020.
[61] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 9.
[62] Ibid., p. 10.

while addressing challenges related to interpretability and robustness. Advancements in hardware, such as the development of specialized AI chips, have played a crucial role in accelerating the training and inference processes of deep learning models, making them more practical and accessible for various applications. Due to the this rapidly increasing processing power of computer systems, in recent times it has been possible to use much larger deep learning networks. From this development, new theories of how to construct and train deep learning, or neural networks have emerged.[63]

2. Typology of AI

The development of AI has made significant progress in terms of what is called weak or narrow AI. This type of AI is applied in specific applications such as strategic games, language translation, autonomous driving and image recognition. It is used as a support for many commercial services such as trip planning, shopper recommendation systems and ad targeting. Narrow AI can also be utilized in important applications in medical diagnosis, education, and scientific research[64], providing valuable insights and augmenting human capabilities in these domains.

Strong or General AI – also referred to as artificial general intelligence – on the other hand, relates to AI systems that exhibit intelligent behaviour on a level which is at least as advanced as humans in the full range of cognitive capacity. Unlike narrow AI, which focuses on specific tasks, general AI aims to possess human-like intelligence and reasoning abilities, enabling it to perform a wide range of cognitive tasks across different domains without the need for specialized programming for each task. Achieving this level of AI would mark a significant milestone in the field of artificial intelligence and have profound implications for society.

However, there is still a wide gap between today's narrow AI and the much more complex general AI – only little progress has been made from the

[63] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 10.
[64] Ibid., p. 7.

expansion of narrow to general AI – and this in spite of the fact that a significant amount of research has been undertaken in this area in recent decades.[65] Developing general AI involves addressing numerous challenges, including creating AI systems capable of learning and reasoning from diverse and limited data, understanding context and nuance in language, and exhibiting common-sense reasoning abilities that humans effortlessly possess. Indeed, among the private sector community, it is generally acknowledged that achieving general AI will still, at the very least, take decades.[66]

For a considerable period of time, individuals have contemplated the potential outcomes of computers surpassing humans in intelligence. Certain experts forecast that an AI with adequate intelligence could be assigned to produce superior and more intelligent systems, which could subsequently be utilized to design systems with even greater intelligence, and so forth. This process may ultimately lead to an "intelligence explosion" or "singularity" where machines swiftly exceed human intelligence. Though, there is no guarantee or assurance that the scenario being discussed will happen in the future.[67]

However, in contrast to many dystopian visions of super-intelligent AI, exceeding human intelligence and abilities, and taking control over humanity, researchers have a more positive view of AI in sense of helpers, assistants, trainers and teammates designed to operate safely and according to our ethical views.[68]

3. Typology of Robots

According to the International Federation of Robotics (IFR), robot types are primarily grouped into industrial and service robots.[69] A third category of robots are social robots. Typically, all three types of robots are physically embodied autonomous agents, interacting and communicating with human beings on an

[65] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 7.
[66] Ibid., p. 10.
[67] Ibid., p. 8.
[68] Ibid.
[69] For more information see: https://ifr.org/.

emotional level.[70] Industrial robots are commonly found in manufacturing settings, where they perform repetitive and precise tasks, often replacing or assisting human workers in hazardous or labor-intensive operations. Service robots, on the other hand, are designed to assist humans in various tasks, such as cleaning robots, medical robots for surgeries, and delivery robots for logistics.

Another interesting concept of robots are cyborgs. These are hybrids of humans and machines.[71] Cyborgs incorporate technology into the human body, augmenting or enhancing their capabilities. This integration of man and machine can take various forms, from simple wearable devices to more advanced implants that directly interface with the human nervous system. Cyborg technology has the potential to revolutionize healthcare, accessibility, and human performance in numerous fields.

The increasing development of machines expanding human abilities raises many ethical and legal questions, e.g., the criminal responsibility of humans equipped with so called brain pacemakers.[72] As technology evolves to the point of enhancing human cognition and decision-making, questions arise about accountability and agency. If individuals use brain pacemakers or other cognitive-enhancing technologies that impact their decision-making processes, how should the legal system address issues of criminal responsibility?

4. Technical basics of AI, Machine Learning and Algorithms

In recent years, AI has been one of the most hyped technology fields, which has, in turn, also made it one of the fastest developing.[73] For example, large technology groups including Baidu and Google made investments between $20

[70] Darling, K. *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects.* In. *We Robot Conference 2012,* University of Miami, 2012, p. 3 ff.; Bekey. G. A. *Current Trends in Robotics: Technology and Ethics*. In. *Robot Ethics - The Ethical and Social Implications of Robotics, Cambridge, Massachusetts: The MIT Press,* 2012, p. 29 ff.
[71] Beck, S. *Grundlegende Fragen zum rechtlichen Umgang mit Robotik.* In. *Juristische Rundschau*, 6/2009, p. 225.
[72] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 597.
[73] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance And the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 38.

and $30 billion in 2016. 90% of the money was spent on research, R&D and deployment, and 10 % on AI acquisitions.[74]

One of the most important technical factors for the development of AI is machine learning. This also forms the basis for a variety of recent advances and commercial applications of AI. State-of-the-art machine learning is a statistical process. The process starts with a set of data, and then attempts are made to derive rules that explain the data. By analysing patterns and correlations within the data, machine learning algorithms can identify underlying relationships and make predictions or decisions based on this knowledge. This specific approach differs from the older, so called 'expert system' approach, where programmers work together with human domain experts in order to learn rules and criteria of decision-making, to translate these rules and criteria into codes or software. This type of system emulates principles used by human experts while, in practice, machine learning is based on statistical methods to find a decision-making procedure.[75] The distinction between expert systems and machine learning reflects two different paradigms in AI development. Expert systems rely on explicit programming and predetermined rules, which can be labour-intensive to develop and maintain, especially when dealing with complex and evolving domains. On the other hand, machine learning offers a more data-driven and flexible approach, allowing AI systems to learn from data and adapt to changing conditions without requiring explicit rule-writing for every possible scenario.

4.1. Technical Aspects of AI

The question of whether responsibility in the form of a specific legal capacity for individual decisions of AI in certain subtasks shows a conceptional proximity to human beings is assessed heterogeneously.[76] Some scholars maintain that autonomous machines are much better (or at least potentially better) suited for their specific tasks. They argue that AI and autonomous systems

[74] Columbus, L. McKinsey's State of Machine Learning And AI, 2017, https://www.forbes.com/sites/louiscolumbus/2017/07/09/mckinseys-state-of-machine-learning-and-ai-2017/?sh=1de4387175b6.
[75] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 8.
[76] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme – Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 146.

can excel in tasks that require precision, speed, and handling vast amounts of data, which may surpass human capabilities in certain fields. Schirrmacher refers to the critical opinions of various authors that machines are often more suitable than humans: "There is something better than humans when it comes to doing business. You just have to make people give more legitimacy and authority to the automated agent."[77] This perspective suggests that in certain contexts, AI and automated agents may outperform humans in decision-making.

It is held that, as a result of their intelligence, autonomously acting machines or software agents can indeed find solutions for complex problems which require a certain degree of rationality. These AI agents can analyse vast datasets, recognize patterns, and derive insights that can aid in making informed decisions. Thus, these agents show an extent of problem-solving and conclusion-drawing capacities, at least remotely similar to the ability of human beings.[78]

However, the behaviour of software agents is based on technical rules, or algorithms. The actionability of these algorithms is generally autonomous and flexible, so that their functionality and behaviour can adapt to changes in their environment. This ability requires appropriate design of the applied mathematical rules and programming, enabling the software agents to respond to dynamic situations and evolving data.

In other words, it means that a computer program is a processing regulation, or an algorithm, consisting of a deterministically programmed sequence of orders, which are formulated into machine codes, and which are able to solve a problem exactly as dictated. Each step of a process is precisely defined. Thus, the algorithm will always deliver the exact same result.[79] This deterministic nature is typical for traditional computer programs with fixed rules and actions.

However, this will not be the case with autonomous, or intelligent software. The ability of problem solving is based on a flexible algorithm. Autonomous

---

[77] Schirrmacher, F. *EGO-Das Spiel des Lebens.* Munich, Karl Blessing Verlag, 2013, p. 148.
[78] Kirn, S. *Integration von Organisation und Informationssystem: Benötigen wir eine Re-Vitalisierung des maschinellen Aufgabenträgers?* Work report, Technische Universität Ilmenau, 1996, pp. 28 ff.
[79] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme – Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 147.

software agents differ from the non-autonomous ones in that they possess learning capabilities.[80] These agents can adapt and improve their performance based on their interactions with data and their environment, allowing them to evolve their decision-making process over time.

Autonomous software agents have independent autonomous behaviour, which neither the programmer, nor the user can predict. With increasing independence, or autonomy of such software agents, the will of the user or programmer becomes increasingly uninfluential. It is an interesting fact that the programmer or user do not know the exact extent of its autonomy. This means that, depending on the extent of the granted autonomy, they will not be able to estimate its degree of learning capabilities. This especially applies to cooperative self-learning software, where the interactions with other agents further complicate the predictability of behaviour. However, this does not necessarily prevent the user or programmer from being liable for the software's actions.[81]

In several decisions, for instance, the German Federal High Court of Justice has held the opinion that the protection of all parties of legal relations must be protected, despite the unknowingness of the machine's actual autonomy.[82] Even though these decisions were dealing with problems in private law, considering such ideas is important in criminal law as well, since these cases may involve criminal actions resulting in loss of money or assets, injuries or even death.

4.2. Human-Machine Teaming

In addition to automation as substitution for a human workforce, there is of course place for the idea of machines that complement human work. In many cases, this concept may be a side effect of AI development, but this is not to exclude the fact that such systems may also be developed specifically to create a human-machine team as well.[83] By combining the specific strengths and

[80] Kirn, S. Kooperierende intelligente Softwareagenten. Wirtschaftsinformatik, 44(1), 2002, pp. 57–60.
[81] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme – Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 148.
[82] BGH, 29.11.1994 – XI ZR 175/93 – NJW 1995, p. 953; BGH, 19.02.2002 – V ZR 17/02 – NJW 2002, pp. 3629, 2631.
[83] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 10.

abilities of humans and AI, human-machine teams are estimated to achieve enhanced performance and efficiency in various tasks and industries.

Indeed, there are many cases where 'teamwork' between human beings and machines is more effective than when either 'goes it alone'. Here the strength of one can compensate for the weakness of the other. Human cognition and intuition can complement the analytical capabilities of AI, and vice versa, leading to better decision-making and problem-solving. An important example of such human-machine teaming is evident in radiology. Taken from a recent study, AI was given images of lymph nodes cells. Its task was to determine whether the cells were cancerous or not. AI alone had an error rate of 7.5 percent, while a human pathologist had an error rate of 3.5 percent. However, the combination of AI and the human pathologist reduced the error rate to 0.5 percent, resulting in an impressive 85 percent reduction in errors.[84] This illustrates how the collaboration between human expertise and AI's data processing capabilities can significantly improve accuracy and outcomes in critical tasks.

Of course, it must be kept in mind that the idea of human-machine teaming brings the interactions between humans and machines closer. Hence, in cases where a person is either physically injured or even killed, this situation will make it more difficult to determine who the criminally liable entity is.

5. Preliminary Thoughts: Historical Background of Criminal Law

Before examining the status of AI in criminal law, it makes sense to look for 'similar' approaches of liability of agents from the past. Studying past approaches and principles for agent liability can offer valuable insights and help inform the development of appropriate legal frameworks for AI responsibility. Research shows that a possibly applicable approach was already created during the Roman Empire.

[84] Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence*. 2016, p. 10, the study is available at: https://arxiv.org/pdf/1606.05718v1.pdf.

Since ancient times, one of the most basic concerns of law has been the clearest possible allocation of liability. In terms of AI, legal discussions bring up many arguments – of which some are mentioned in this thesis – that are both not new, and somehow also recall well-known examples of antique principles for the allocation of liability within the use of work animals or slaves. The similarities between the debates surrounding AI liability and historical cases involving work animals and slaves lie in the question of determining responsibility when an agent, whether human or non-human, causes harm or damage through their actions. Who, in ancient Rome for instance, was liable if a workhorse passed by and caused damage to property, or if a slave, working as an engineer, made mistakes while constructing a bridge, which subsequently made the bridge collapse and kill several people?[85]

5.1. Robot-Law in the Light of Roman Law

The historical context of liability concerning slaves provides an interesting perspective on the allocation of responsibility for the actions of agents. Whenever a slave committed a misdemeanour, his holder's responsibility was limited to civil liability. Even though slaves were entities with legal competence, they had no legal capacity, which prevented them from having legal obligations.[86] The consequence was that the slave owner was the only one to be liable for his slave's actions. However, some variation in the cases did exist, allowing two possible forms of liability. First, if the owner ordered the slave to commit an unlawful act, then the former was to be held liable, because he was responsible for the slave's action and, in essence, used the slave as a means through which harm or damage could be caused. The second variation concerned where the owner allowed the unlawful act to happen through negligence. For this scenario, the owner had to know about the slave's plan to commit the particular unlawful act and then not prevent him from doing so. If this was proven, the owner was held liable as the perpetrator of the unlawful act. He was then compelled to compensate any damages the slave had caused to other person's

[85] Gless, S./Seelmann, K. *Intelligente Agenten und das Recht – Verantwortungszuschreibung in Antike und Moderne.* In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 11.
[86] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 97.

legally protected rights.[87] This historical approach to liability reflects some principles of causation and responsibility that are still relevant in contemporary legal systems.

The obligation to assume responsibility for slaves and work animals was entirely different to liability where the person himself is at fault. According to Roman law, animals could not be held liable since animals do not possess intellect or the capacity for intention. As a result, according to Roman lawyers, animals could not carry out unlawful acts. This implies that the holder of the animal could not be held responsible for misconduct or for any similar, liability-inducing deed by the animal. Instead, only the autonomous failure of the animal was sanctioned. To exclude the involvement of the holder, the harmed, or third person from the occurrence of the damage, Roman jurists demanded that the animal caused the damage contrary to its tame nature.[88] That means that the animal's owner was not liable for the animal's actions if the animal was provoked or incited by someone else, for example, by causing it physical pain or intentionally irritating the animal. In such cases, the responsibility for the animal's actions would not rest on the owner, but rather on the person who provoked the animal.

In a later development, during the high classical era of Roman law, Celsus argued for making exceptions to the so-called noxal surrender,[89] when a slave caused damage to property or harm to another person. According to Celsus, noxal surrender should not be a possibility for a slave holder where he was liable for the slave-caused damages by ordering the slave to commit a harmful deed or, at least, contrary to his duty, or failed to conduct measures to prevent the slave from doing so. In terms of liability, this idea had a relieving effect on slaves, since vetoing could result in being killed by their owners.[90]

---

[87] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 97.
[88] Ibid., p. 107.
[89] For a definition see e.g., https://en.wikipedia.org/wiki/Noxal_surrender.
[90] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 108.

However, Celsus's opinion did not become the prevailing view. Instead, the opinion of the high classical era jurist Julian became the prevailing view. He saw that the owner's liability would not relieve the slave, but rather put the holder under additional liability. What the opinions of Celsus and Julians had in common was that slaves are not objects, but independent human perpetrators. Contrary to Celsus, Julian's opinion did not forgive the slaves deeds solely based on their subordinate position, however.[91] Thus, the slave holder had to take responsibility, if he did not contribute at all to the slave's deeds and could only be sued based on the noxal surrender only, just as holders were liable for damages caused by their animals. In both cases it was not about a potential hazard created by the holder, but rather that the holder had to assume responsibility for those damages not attributable to the holder only because he is the sole person to whom liability can be allocated. In particular this was applicable to cases where the owner had to compensate damages caused by the slave before the owner purchased him.[92]

Major differences in the dimensions of liability can be seen when a slave holder employed his own slaves, and when he rented slaves from other owners. Whenever the holder's own slave damaged the legally protected rights of other people, the owner was held liable according to the noxal liability, which meant that he had to give his slave to the person whose legally protected rights had been infringed. Important here is the fact that noxal liability was relevant, and that the owner was obliged to surrender the slave when, for example, the damage was caused by the business owner's own slave.[93] If the business owner rented slaves from other owners, he was held liable in a form of strict liability. The difference was that the latter case did not lead to noxal surrender as a consequence for the owner, who rented his slave out to the business owner. This led to legal circumstances which, at first glance, seemed imbalanced. Important factor here was that the principal decided to use external slaves and thus, created a risk potential. However, putting the principal under a stricter liability regime within the deeds of a rented slave followed a certain logic: owners or principals knew

[91] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 110.
[92] Ibid., p. 110.
[93] Ibid., p. 112.

their own slaves with their competences, characteristics, and reliability much better than the foreign, or rented slaves. The Jurist Ulpian from the late classical period justifies the difference in his commentary on the lawsuit for property damage by explaining that a shipowner is forced to assess the suitability of foreign slaves, while he deserves consideration in using his own slaves as they are, according to their character traits.[94]

After evaluating the considerations surrounding Roman slavery law and its potential applicability to AI responsibility, the question arises as to whether the responsibility of AI should follow the ideas of Roman slavery law. While there are some similarities in the considerations of control, knowledge, and risk assessment in determining liability for actions carried out by AI systems, applying a liability approach similar to that of noxal surrender is not a contemporary solution for today's standards.

In the context of AI, similar considerations of control, knowledge, and risk assessment are relevant in determining liability for actions carried out by AI systems. The basic notion that the owner of the machines is liable for the deeds of the machine, just as slave owners were liable for their slaves' acts seems correct at first glance. However, despite the similarities to strict liability, this approach is not contemporary for today's standards. This is because, if AI is involved, attributing liability cannot be performed in such a simple manner, since there are too many factors influenced by other relevant parties which have their own impact in the value chain of an AI product. Especially the noxal surrender seems like a 'primitive' solution, for the reason alone that mostly nobody will accept a malfunctioning machine, such as an autonomous vehicle, as recompense, rather the person whose property or other legally protected rights are damaged will most probably demand monetary compensation for the damages.

The only concept taken from Roman slavery law which seems most acceptable is the principle of socialisation, namely, if the use of autonomous, intelligent agents is socially accepted and desired, the individual user may not

[94] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht.* 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 113.

be held unlimitedly liable for its failure or malfunctions.[95] This principle aligns with modern notions of shared responsibility and collective action, recognizing that the widespread adoption of intelligent agents' benefits society as a whole. If society takes profit from the use of such agents, the consequences must also be socialized, for example, by mandatory insurance which minimises the risks of the users' or owners' personal liability.[96] By implementing mandatory insurance, it creates a safety net for potential damages caused by these agents and encourages responsible usage among users and owners.

6. Applying Product Liability on AI

As intelligent machines, equipped with general AI and fully automated AI systems with independent learning capabilities continue to evolve, the judiciary will be confronted with increasingly complicated issues of liability. In the US, litigation concerning harms caused by automated, and AI systems, are already part of courts' concerns.[97] Certainly, as technology is developing into the direction of an increasingly higher degree of autonomy, law regimes are being pushed towards their limits as the application of new technologies to traditional theories of liability becomes increasingly confusing.[98]

In light of these challenges, legal experts and policymakers are recognizing the need for proactive measures to address the unique challenges posed by AI systems. The current legal frameworks may not always adequately address the unique characteristics of AI systems, including their autonomy, learning capabilities, and complex decision-making processes. This creates uncertainties and challenges in applying existing liability principles to cases involving AI.

How the judiciary applies the current, existing legal frameworks to remedy harms will probably play a major role in governing AI developing towards

[95] Harke, J. D. *Sklavenhalterhaftung in Rom*. In. *Intelligente Agenten und das Recht*. 1st ed. vol. 9, Baden-Baden, Nomos, 2016, p. 113.
[96] Ibid., p. 117.
[97] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 54.
[98] Ibid., p. 55.

increasing autonomy. Achieving a balance between encouraging innovation and safeguarding against potential harms will be one of the major tasks for lawmakers and judges alike.

6.1. Product Liability

Product liability might be an example for a probably very acceptable form of legal framework for assigning responsibility for damages, or injuries caused by AI systems. Product liability typically involves claims concerning defects in manufacturing, design, information or a failure to warn. Errors such as flaws in programming, the use of low-quality data during the machine learning process, or failing to warn consumers, or users etc. of probable dangerous consequences, can make AI fall within the scope of product liability.[99] However, product liability is based on the idea of identifiable faults. Applying product liability to AI systems can be challenging due to the nature of AI's complex decision-making processes. Unlike traditional products with tangible defects, AI systems' "defects" may not be easily identifiable or traceable to a specific flaw in the traditional sense. As an example of a crashing plane, due to a fault in its autopilot system, the manufacturer of the system is likely to be held liable.[100] Another example is a situation where an AI-powered medical device makes an erroneous diagnosis. The issue might not be caused by a single identifiable programming bug but could be the result of complex interactions within the AI's neural networks. US-courts have already been confronted with a case concerning the question of whether the manufacturer may be held liable in cases where the presence of a defect is evident, but cannot be identified. A suspected software defect in vehicles produced by Toyota caused a sudden acceleration that was uncontrollable by the driver in a particular case.[101] Here engineers were unable to identify the exact flaw which led to the acceleration.[102] Even though the plaintiffs failed to prove the specific flaw which led to the defect, the court found

[99] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 58.
[100] Ibid.
[101] Ibid., p. 58; see Fleck v. General Motors LLC, 14-CV-8176.
[102] See Fleck v. General Motors LLC, 14-CV-8176.

sufficient evidence that the accident was likely caused by the car itself and not the driver.[103]

Significantly more difficult to consider is the question whether product liability suffices to correct harm in cases where it may not be sufficiently concluded that a specific defect contributed to the injury caused by the AI system.[104] However, when an AI machine diverges from its programmed behaviour in a sense of true autonomy, the agency principles may lose relevance.[105] Considering this, the question arises as to how to determine autonomy as defective in the meaning of the law and how such risks are to be managed. Who shall bare the loss in cases where there is an injury but no apparent defect, or failure that stems from a human contribution? Product liability in its traditional concepts is very likely to fail in such cases, as manufacturing defects cannot always be found.[106]

In cases without direct evidence of fault, common law traditionally goes towards the doctrine of res ipsa loquitur. However, the idea that negligence is concluded by the nature of harm, does not answer the liability questions for fully autonomous AI systems.[107] Under the concept of res ipsa loquitur, defendants have the possibility to negate the inference of the necessary elements of duty of care, breach and causation by evidentially showing that their conduct was not negligent. Furthermore, it is based on the conclusion that someone is at fault, but if the specific harm is unexplainable, and untraceable, the prerequisites of res ipsa loquitur cannot be fulfilled.[108]

In terms of autonomous driving, some scholars state that AI systems do not pose any specific questions concerning products liability and its corresponding

---

[103] See In re: Toyota Motor Corp. Unintended Acceleration Mktg., Sales Practices, & Prod. Liab. Litig., 978 F. Supp. 2d 1101 (C.D. Cal. 2013).
[104] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance And the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 59.
[105] Vladeck, D. C. *Machines without principals: Liability rules and artificial intelligence.* In. *Washington law review*. 89(1), 2014, p. 142.
[106] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance And the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 59.
[107] Ibid.
[108] Ibid.

duties.[109] If AI systems cause any damage, it depends on who induced the attributable causes which led to the occurrence of the damage.[110] Thus, all producers' duties and obligations are easily applicable to the production of AI. In this regard, producers are, as usual, obliged to consider all sources of information to avoid risks arising from their products.[111] The significant factor here is whether the producer should have known about the error when bringing the product to the market.[112] Risks arising from a product which occur after their release do not lead to liability of the producer for design faults.[113] Thus, only subsequent obligations within product monitoring may be considered here.[114] Also the IT-sector is of the opinion that, despite the complexity of IT-systems (which covers AI as well), there is no alternative to fixing safety issues immediately, before the product is launched on the market.[115] Further, producers are obliged to inform customers about risks from using the specific product. This also counts for AI systems. Costumers must be informed about the correct use of the system. However, if such systems are made for professionals, the producer's obligation to instruct the professional costumers may be reduced significantly.[116] In cases where AI systems are designed and marketed for professional users, the level of expertise and training of these users might impact the extent of the producer's responsibility as well.

As described in Part 6 Chapter 2 (6.2.) producers of AI systems are obliged to monitor their AI-products after their launch on the market. The reason why a product-monitoring obligation is acknowledged is that the producer's duty of

---

[109] See Spindler, G. *Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen.* In. *Computer und Recht*. 12/2015, 2015, 766 – 776.
[110] See Günther, J./Böglmüller, M. *Künstliche Intelligenz in der Arbeitswelt.* In. *Betriebs Berater.* 2017, 53-58.
[111] BGH. 12.11.1991 – VI ZR 7/91, NJW 1992, 560; BGH, 17.10.1989 – VI ZR 258/88, NJW 1992, 560; Foerste, U./Westphalen, F. Graf v. *Produkthaftungshandbuch.* 3rd ed. Munich, C.H. Beck, 2012, § 24 recital. 378.
[112] *Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen.* In. *Computer und Recht*. 12/2015, 2015, 766 – 776.
[113] Jänich, V. M. et al., *Rechtsprobleme des Autonomen Fahrens. Neue Zeitschrift für Verkehrsrecht*, 7/2015, p. 317.
[114] Söbbing, T. *Fundamentale Rechtsfragen künstlicher Intelligenz.* Frankfurt am Main: Fachmedien Recht und Wirtschaft. 2019, p. 139.
[115] Bartsch, M. *Computerviren und Produkthaftung.* In. *Computer und Recht.* 11/2000, p 722 ff.
[116] BGH, 14.05.1996 - VI ZR 158/95, NJW 1996, 2224.

care is related to a certain point of time, and based on historical risk assessments.[117] The extent of this obligation depends on several factors, such as the possible damages, the possibility of the damage occurring, and how far monitoring is economically reasonable for the producer.[118] The obligation of monitoring is less pronounced towards products which have existed for a long time, and those which are in a high quantity. At the same time, the obligation becomes more intensive towards newly developed products of high complexity and damage potential.[119] In this context, AI systems may be treated as most other IT-systems, as it is historically well known that some programming failures, or so-called bugs, are unavoidable. As a result, producers are obliged to continue monitoring AI-systems very carefully after the release.[120] The purpose of the monitoring obligation is to gather as much data about the product as possible about possible risks. However, sources of such information may not solely be the manufacturer's own products.[121] Indeed, sources may also be the same or at least similar products of competing producers, as far as this information is legally available to them.[122] Producers are already obliged to provide information if their product is seriously suspected to be dangerous;[123] for example, in the case of autonomous vehicles, the obligations associated with monitoring and reporting potential risks are even more stringent, considering the high stakes and potential impact on public safety.[124]

6.2. Criminal Product Liability

Criminal liability might also emerge from product liability. For this to be the case, first, it needs to be examined whether criminal product liability exists at all. The question of the existence of this form of liability arises whenever a supposedly safe product is introduced to the market, and causes some form of

[117] BGH, 17.03.1981 - VI ZR 286/78, NJW 1981, 1606, 1607 ff.
[118] Wagner, G. In. Säcker et al. *Münchener Kommentar zum Bürgerlichen Gesetzbuch* 7th ed., vol. 6, München, C.H. Beck, 2017, BGB § 823 recital 836-837.
[119] BGH, 09.12.1986 - VI ZR 65/86, NJW-RR 1995, 342 p. 343.
[120] Foerste, U./Westphalen, F. Graf v. *Produkthaftungshandbuch.* 3rd ed. Munich, C.H. Beck, 2012, § 24 recital. 174.ff.
[121] Wagner, G. In. Säcker et al. *Münchener Kommentar zum Bürgerlichen Gesetzbuch* 7th ed., vol. 6, München, C.H. Beck, 2017, BGB § 823 recital 839.
[122] Ibid., § 823 recital 375.
[123] BGH, 17.03.1981 – VI ZR 191/79, BGHZ 80, 186, 192.
[124] Söbbing, T. *Fundamentale Rechtsfragen künstlicher Intelligenz.* Frankfurt am Main: Fachmedien Recht und Wirtschaft. 2019, p. 140.

harm to the legal interests of a person. The history of product liability knows some spectacular cases, where criminal product liability was discussed, such as the Thalidomide case, a sedative sold between 1957 and 1961. It was primarily prescribed to pregnant women and led to severe disturbances in the growth of foetuses. Even after several scientists and medical doctors pointed out a possible connection between the consumption of Thalidomide and malformations, it remained on the market. This case serves as an example of the challenges of product liability, as it highlights the delicate balance between therapeutic benefits and potential risks that may go unnoticed.[125] As with medicine and many other technologies, AI raises many legal questions when a product using it is launched after many years of research and development, and then turns out to cause harm to human beings.

The most critical aspects in this regard are the problems that stem from legislators who permit these newly developed technologies to be placed on the market. In the main, producers face criminal liability if it had been clear to them saw, or at least should have been clear to them, that their product is dangerous before it was launched on the market. However, it must be considered that innovative and new technologies are always dangerous to a certain, even if it is only a minimal, degree. The risk of innovation can never be excluded entirely.[126] This was the case with the new Thalidomide, and so will it be with AI.

According to the 'Lederspray-Entscheidung'[127] ('Leather Spray decision'), managing directors of companies are legally responsible for ensuring that consumers of the products, produced and distributed by the companies were not exposed to health risks that could arise from the intended use of these products due to their nature. This decision sets a precedent for holding corporate leaders accountable for the safety of their products and places a legal obligation on them to prevent harm to consumers. Anyone who places hazardous consumer products on the market is obligated to prevent harm, and if they negligently fail to fulfil this duty, they must be held criminally liable for any resulting damages.[128] The

[125] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 54.
[126] Ibid., p. 55.
[127] BGH, 06.07.1990 - 2 StR 549/89; BGHSt 37, 106.
[128] BGH, 06.07.1990 - 2 StR 549/89, recital 34.

criminal court derives this obligation from the duty to ensure public safety under civil law or, to be more precise: the product monitoring obligation. According to the criminal court, there are good reasons why the same civil obligations, which are decisive for product liability, also form a basis for criminal liability, as the obligation of compensation for physical harm, caused by faulty products is seen as a case of torts law. The producer in the 'Leather Spray decision' was in a position of guarantor (in German law) because of previous dangerous deeds, contrary to their duties.[129] The court's recognition of the producer as a guarantor emphasizes the responsibility and duty of companies to take proactive measures to avoid harm to consumers and maintain a high standard of product safety. The 'Leather Spray decision' says that the company that has created a dangerous situation for others -in this particular case by selling a dangerous product to consumers through negligent behaviour is obligated to prevent the impending harm. This applies at least when the behaviour makes the risk of harm appear obvious and the negligence consists precisely in the violation of a duty designed to protect the endangered legal interest.[130] Within this framework, there is also criminal liability for the production and distribution of faulty products. Producers which negligently cause a danger to the consumers of such products by putting them on the market must generally be held responsible for ensuring that this danger does not result in corresponding harm. This applies to the production and distribution of consumer goods that are designed in such a way that their intended use by consumers -contrary to their reasonable expectations- poses a risk to the consumers health. In this regard, liability applies not only to those who cause harm through an active deed but also to those who fail to prevent the impending harm.[131]

Further, a Swiss criminal court developed a very similar legislation for criminal products liability: carelessness is a breach of duties if the perpetrator, based on the circumstances and his knowledge and skills, should have known or recognised the possible physical risks his deed might cause the victims, while he crosses the limits of permitted risk.[132] This means, first, that the requirements of

[129] BGH, 06.07.1990 - 2 StR 549/89, recital 35.
[130] BGH, 06.07.1990 - 2 StR 549/89, recital 36; BGHSt 37, 106.
[131] BGH, 06.07.1990 - 2 StR 549/89, recital 37.
[132] BGE 121 IV 10, p. 14.

duty of care need to be measured for every case individually.[133] Further, the duty of care, and the breach of it needs to be apparent and avoidable.[134] This approach emphasizes the importance of assessing each case individually and considering the specific circumstances, knowledge, and skills of the perpetrator when determining criminal product liability.

These considerations pose the question when exactly criminal products liability applies and what the exact duties of care concerning innovative products are.[135]

To measure the degree of the duty of care, in many legislations criminal law refers to other laws, directives and regulations. However, some of these refer to duties of producers *while* products are brought onto the market and others to duties *after* the product has been brought onto the market. Producers must bring fault and error-free products to the market, measured on the current state of science and technology. Thus, they must manufacture these products in accordance with standards and test them prior to their launch, to be able to detect and avoid risks for human beings. Producers generally have to subject their - potentially hazardous - product to an authorisation process[136] while complying with the relevant regulations. However, an authorisation does not exempt the producer from liability for errors that could have been avoided according to the state of science and technology.[137] The authorization process still serves a highly necessary step to ensure the safety of products before they reach the market. However, the liability risk remaining with producers is partially viewed critically because producers never achieve legal certainty. In addition, there is a risk that, after damage has occurred, the judge -influenced by the damage that has occurred- is more likely to make a retrospective judgment based on the state of science and technology, which would require a careful producer to have known

---

[133] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 56.
[134] Spitz, P. *Strafrechtliche Produkthaftung – Übertragbarkeit zivilrechtlicher Betrachtungswesen? Unter besonderer Berücksichtigung der Organisationshaftung in Straf- und Zivilrecht.*Basel, Basel, Helbing & Lichtenhahn, 2001, p. 68 ff.
[135] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 56.
[136] E.g. German Medicinal Products Act and the Product Safety Act.
[137] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 56.

that their product was dangerous, ultimately shifting the innovation risk back to the producer. [138]

Producers further need to observe their products in action, even after they have been released as safe to use. Liability may be imposed on producers retrospectively for harmful products if they are left on the market nevertheless.[139] This means that producers have an obligation to keep track over their sold products, and take measures to avoid damages when defects are discovered subsequently. In this case, the producer is not liable for producing or delivering such products, but for refraining from subsequent monitoring, information of the costumers, and recalling the product if necessary.[140] The obligation to observe the product is a general duty of care and is therefore also relevant for establishing a standard of negligence liability. This means that the duty of care extends beyond the initial product launch and continues throughout the product's lifecycle, emphasizing the need for ongoing oversight and prompt action to address any identified risks. The careful producer must therefore observe his product and react to any dangers that may be identified. If he fails to do so, he is liable.[141]

7. Intermediate Conclusion

Concerning AI, the abovementioned rules apply in the same way. Producers of autonomous cars have to test the products according to the current state of science and technology to avoid liability in any form, if their product causes any damage or harm to the legal interests of persons. If an autonomous car is brought to the market, producers are obliged to monitor their products to remove errors, for example, by a firmware update if possible, or by recalling their faulty product to ensure the safety and reliability of these technologies. This is especially crucial for autonomous cars, as their operation directly impacts public safety and human lives. If producers avoid obeying their duty of care and duty of

[138] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 56.
[139] Ibid., p. 59.
[140] BGE 121 IV 10, 14; BGH, 09.12.1986 - VI ZR 65/86 - NJW-RR 1995, 342.
[141] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 60.

monitoring, criminal liability is possible. However, this approach may be inapplicable to strong AI, as these systems are self-learning systems, probably at least as intelligent as human beings. Such systems cannot be connected to their producers for an unlimited period. This leads to the problem that testing in accordance with the current state of science and technology needs defined standards. The traditional framework of product monitoring and testing may need to be re-evaluated and adapted to accommodate the complexities of strong AI systems. Currently, there are no established standards for highly innovative or ground-breaking technologies like strong AI, which is due to the fact that Strong AI does not yet exist. Testing autonomous cars however, appears simple compared to the standards that need to be set for testing strong AI.

If criminal liability is to be 'inspired' by civil strict liability, it needs to be considered that, in certain cases, such approaches cannot be taken as role models to make someone liable for faulty products. While civil strict liability may serve as a basis for determining certain aspects of criminal liability, it is essential to acknowledge the distinctions between civil and criminal law, as the goals and considerations differ significantly in the two fields. The basic principle holds that anyone who establishes a risk shall be responsible for taking all measures to avoid damaging third parties, if the risk exceeds the basic life risk.[142] However, producers are not obliged to avoid risks which are unforeseeable according to the current state of science and technology. Imposing an obligation on producers to avoid unforeseeable risks that are beyond the current state of science and technology could create significant barriers to innovation and hamper technological advancements.[143] Criminal law considers unknown risks as inherent to the basic risks of life[144], acknowledging that absolute certainty and foresight are not always achievable and that risks are an inherent part of human life and technological progress.

[142] Seelmann, K. *Nichtstun als Straftat – Bedeutungszuwachs und Problempotential der unechten Unterlassungsdelikte.* In. *Schweizerische Zeitschrift für Strafrecht*. 3/2007, p. 268.
[143] Seelmann, K. *Privatrechtlich begründete Garantenpflichten?* In. Schmidt K. *Vielfalt des Rechts – Einheit der Rechtsordnung? Hamburger Rechtsstudien,* iss. 85, 1994, Berlin, Duncker und Humblot, 1994, p. 91.
[144] Zech, H. *Gefährdungshaftung und neue Technologien.* In. *JuristenZeitung.* 68(1), 2013, p. 23.

In conclusion, careful companies which follow their duties of care by carefully testing their products according to the state of science and technology, will most probably not be held liable for bringing products onto the market where the fault cannot be detected after careful testing. Therefore, meeting the obligations of conducting rigorous testing based on the current state of knowledge and technology helps establish a strong defence against liability claims for unforeseeable faults or defects that were not detectable at the time of product launch. According to many jurisdictions, producers fulfil their duties by respecting national and international licencing regulations, search for risks according to the current state of science and technology and inform customers about apparent risks.[145] However, a producer will most likely be liable for damages, if it`s product is successfully tested according to all relevant regulations, but is aware of the fact that the product still has errors which will likely cause damages in future.[146] In cases where producers are aware of potential risks or defects but still choose to release the product without addressing these issues, they may be held accountable for any damages that result from their negligent actions, as they knowingly put consumers at risk despite being aware of the product's flaws.

8. The Defendant as an Alternative Legal Entity?

The perception of AI systems and the legal system which underlies it is a significant prerequisite for resolving issues of liability for harms caused by AI,[147] especially in terms of criminal liability. Autonomous machines, no matter how far developed they are, share one common critical feature in assessing liability: every machine function and decision made work in ways that can be tracked directly back to the design, programming and knowledge that a human being implemented in the system.[148] The traceability of AI system actions to human

[145] Fellmann, W. /von Büren-von Moos, G. *Grundriss der Produkthaftpflicht.* Bern, Stämpfli Verlag, 1993, p. 212 ff.
[146] Gless, S. *Strafrechtliche Produkthaftung.* In. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, p. 59.
[147] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 55.
[148] Vladeck, D. C. *Machines without principals: Liability rules and artificial intelligence.* In. *Washington law review*. 89(1), 2014, p. 120.

design and programming serves as a foundational aspect in understanding and attributing responsibility for AI-related harms, as it establishes a clear link between human actions and the system's behaviour. This link becomes crucial in determining liability and accountability for the consequences of AI decisions. If an autonomous system's decision bears clear signs of human involvement, it may be possible to establish fault using common law principles of liability.[149] By applying common law principles of liability, such as negligence or product liability, courts may hold human designers, programmers, or manufacturers responsible for the actions and decisions of AI systems, especially when there is evidence of human involvement in the decision-making process. Under the principal-agent concept, AI systems may be considered an agent of the respective manufacturer, or any other entity that may be held responsible for the damages caused by its machine.[150] The principal-agent relationship recognizes that the actions of an AI system can be attributed to its creators or operators etc., treating the AI system as an extension of the responsible entity.

As mentioned above, AI systems are developing towards increasing autonomy. Thus, existing regulations concerning liability are likely to become insufficient in terms of attribution of liability for any harm caused. Completely autonomous AI systems able to perform tasks based on their own analysis, to a degree of autonomy that no human intervention is required, brings the principal-agent concept into question, leading to uncertainty in assigning responsibility for the actions and decisions of AI systems. The evolving autonomy of AI systems challenges the notion of human control and intervention, making it difficult to apply existing liability models. Here the question arises whether an AI system equipped with such a high degree of autonomy may still be considered as an agent of the principal. Does such a machine break the connection to the manufacturer or creator of the algorithms to such an extent that it can no longer be seen as the agent of the principal? If this is the case, then to whom or what should the liability be attributed? It is probable that faulty data used within the machine learning process, or the programmer failing to forecast possible

[149] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 55.
[150] Ibid.

outcomes may change the conditions concerning liability to the production chain. Yet the question remains as concerns cases where the fault is unclear and cannot be determined, or where it is the result of unforeseen harm.[151] Important situations will be in which robots cause damages, either by injuring or even by killing human beings. A variety of individuals may be considered as perpetrator, for example, the producer, the programmer of the software, the seller, or indeed the user of the robot (which is especially significant in autonomous driving).[152] This is not an exhaustive list of possible perpetrators, as intermediaries, carriers or any other positions 'between' these could come into question.

According to the current technological development of AI, robots can only gain relevance in criminal law as objects or instruments to commit a criminal act.[153] The current state of AI technology limits the direct legal culpability of AI systems in criminal law, as they can be viewed primarily as tools or instruments used to carry out criminal acts, rather than autonomous agents with intent and free will. Possible scenarios raising the question of liability are such as, for example, autonomous cars killing a pedestrian. In this case, liability for negligent physical injury of the producer and/or the user comes into question. Then the violation of possibly existing duties of care, especially the relevance of their breach for the outcome needs to undergo a legal review.[154] Further, it needs to be clarified whether one of the relevant parties has a guarantor position. The complexity of the particular robot causes significant problems for criminal law in connection with the casual link analysis, however. Especially criminal law requires indubitable determination of the deed. In Germany, the question concerning the legal culpability of robots in the light of brain research, and it's astonishing results for the doctrine of the free will of human beings is an interesting aspect.[155] As jurisprudence handles the free will of humans as a required fiction,

---

[151] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 56.
[152] Beck, S. *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 140.
[153] Hanisch, J. *Haftung für Automaten.* In. *Internationale Göttinger Reihe.* Vol. 19, Göttingen, J.-P. Cuvillier, 2010, p. 13.
[154] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 605.
[155] Ibid. For more on the debate of the freedom of will in German criminal law see: Fischer, T. *Strafgesetzbuch mit Nebengesetzen.* Munich, C.H. Beck, 2016, § 20 StGB, recital 3, with

basically a transfer of this fiction to robots seems possible.[156] With the staggering results of brain research, conducted at the beginning of the 21st century, German criminal law, for example, went back to the position, that freedom of will must be a necessary fiction.[157] New technologies in brain research gave scientists new insights into the functioning of the human brain. The insights showed that the driving forces of our decision-making and will are not the results of an immaterial, personal 'being' in an immaterial exchange with the natural and social surroundings, but the outcome of dynamic chemical and physical processes inside the central nervous system, making determined reactions based on the individual neuronal interconnections.[158] Thus, the question arises that, if freedom of will is simply a fiction, why should it not be possible to simply extend it to robots?[159] There is no particular reason why this should not be the case when brain research proves that no metaphysical, but rather chemical and physical processes, determine our actions just like modern processors and software. The philosopher Andreas Matthias has even called for a special criminal law for robots. Matthias asks for the 'legal emancipation'[160] of machines and for making them accountable in criminal law.[161] At this point it is important to mention that, in contrast to Germany, in certain countries such as, for example, the Czech Republic[162] and Switzerland,[163] companies may be held criminally liable. These nations share the fact that they have already broken with the tradition of anthropocentric criminal law by no longer reserving criminal liability exclusively for natural persons.

---

further extensive references.

[156] Hilgendorf, E. *Können Roboter Schuldhaft handeln?* In. Beck, S. *Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs.* Baden-Baden, Nomos, 2012, pp. 129 ff.

[157] Hilgendorf, E. *Können Roboter Schuldhaft handeln?* In. Beck, S. *Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs.* Baden-Baden, Nomos, 2012, pp. 129 ff.

[158] Fischer, T. *Strafgesetzbuch mit Nebengesetzen. Munich.* C.H. Beck, 2016, Vor § 13 StGB, recital 9.

[159] Hilgendorf, E. *Können Roboter Schuldhaft handeln?* In. Beck, S. *Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs.* Baden-Baden, Nomos, 2012, pp. 129 ff.

[160] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 605.

[161] Matthias, A. *Automaten als Träger von Rechten*. 2nd ed. Berlin. Logos Verlag Berlin, 2010, p. 249.

[162] Act No 418/2011 Coll (Czech Criminal Liability Act).

[163] Art. 102 StGB (Swiss).

Thus, the criminal responsibility of AI would be easier to justify in Switzerland and in Czech Republic than in Germany.[164]

8.1. Personhood

In the 1990s a number of scholars questioned whether AI systems should receive legal personhood.[165] Several alternative ways to determine liability based on the agency of AI systems have been developed around the different concepts of legal identity.[166] While the concept of AI receiving legal personhood remains a subject of ongoing debate, scholars believed that this issue is only theoretical since the respective technology that justifies legal review of the matter was not existent back then. Though, recent developments in AI systems with a higher grade of autonomy may soon create the need for proper legal review Although the idea of treating Al as a person may appear unconventional, the existence of juridical entities recognized by law as possessing the rights and obligations of natural persons demonstrates that conferring the legal status of a person to machines is not entirely implausible.[167]

The issue of according legal identity to AI systems will be even more complicated by the differences between the various systems and technologies. What is clear is that all of these systems are shifting towards increasing autonomy, leading to the need to reassess the capacity of the existing, traditional regimes of liability. Litigation is likely to continue being driven by the belief that principals, including developers, manufacturers, and owners, bear direct responsibility for the Al they choose to implement until the judiciary is compelled by technology to embrace new legal identity frameworks for AI systems.[168]

---

[164] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, p. 605.
[165] See Solum, L. B. *Legal Personhood for Artificial Intelligences.* In. *North Carolina Law Review*. 70(4), 1992. pp. 1231 ff.
[166] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 56.
[167] Ibid.
[168] Jackson, B. W. *Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems.* In. *Santa Clara High Technology Law Journal,* 35(4), 2019, p. 57.

According some type of personhood to AI systems may open up the possibility of making these systems both civilly and criminally liable for their actions. However, currently, robots cannot be sued and still, in the instance of a liable robot being a defendant, compensation would still be required at the corporate level. Obligations to pay compensation might also be passed to the owner of the system.

8.2. Philosophical Background of Personhood

In the following argument, we will assume that robots can acquire self-awareness, emotions, empathy, and morals, and that their freedom may not necessarily appear inferior to human beings. Human beings make their supposed free decisions based on an incomprehensible structure of influences from genetics, education, social environment and feelings. Yet these decisions are supposed to be free. In this regard, the question arises where the border lies between machine and human beings as traditional subjects to criminal law: is it enough for human beings, if machines are able to process data from their environment and compare them to existing patterns and to act accordingly, to accept them as responsible 'persons'?[169]

According to the German law philosophy, which is heavily influenced by Immanuel Kant and John Locke, personhood describes an individual who is able to develop an intellectual relationship to himself.[170] Locke assumes that the prerequisite of personhood is that somebody as an intelligent agent is capable of law, and happiness and misery: "this personality extends itself beyond present existence to what is past, only by consciousness, – whereby it becomes concerned and accountable; owns and imputes to itself past actions, just upon the same ground and for the same reason as it does the present."[171] The ability of self-reflection, according to Kant, allows persons to be autonomous. Kant claims that, since humans are self-aware, they need to be tamed by laws which they create for themselves due to their sanity. Due to their sanity, every sane being can say that it could have refrained from committing an unlawful act. Therefore,

[169] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, p. 568.
[170] Ibid., p. 569.
[171] Locke, J. *An Essay Concerning Human Understanding.* 1690, p. 331 (§ XVII Nr. 26).

if a self-aware person in his existential freedom, cannot realise that he, himself is responsible, then other people are entitled to attribute the responsibility to that person as well.[172]

Of course, the above-mentioned considerations did not have any reference to AI or modern robots. However, it is possible to apply these considerations to modern robots and state that AI lacks certain significant elements of personhood. Even though modern AI is able to learn, and even make unforeseeable decisions, it cannot be aware of its freedom, nor can it see itself as having rights and obligations in our society. Intelligent agents can indeed perform certain tasks automatically, but even if they are capable of learning, they ultimately follow the options predetermined by programming, and can provide no justification for its behaviour. Therefore, we do not perceive AI as being free and cannot make it responsible for damages.[173]

8.3. AI in the Light of Philosophy and Science

The discussions concerning the legal capacities of AI systems, or autonomous software agents, share strong similarities with the discussions of philosophers, scientists, and especially neurophysiologists and theologists, about the question of whether human beings have a free will.[174] The latest neuroscientific experiments confirm the view of the conception that our brains, which are subject to scientific laws cannot dictate our actions according to some instance outside of these laws. Everything is subject to the laws of chemistry and physics.[175]

Philosophy defines the freedom of will as the freedom to choose and to act.[176] According to Kant, the only thinkable source of autonomy is the subjective will, if the will is ruled by sanity.[177] In terms of the learning capacities of AI, Kant's

---

[172] Kant, I. *Kritik an der Vernunft.* In. Weischedel, W. *Kant, Werke in Zehn Bänden.* vol. 6, Darmstadt, Wissenschaftliche Buchgesellschaft WBG, 1975, p. 223.
[173] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, p. 570.
[174] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 142.
[175] Hawking, S./Mlodinow, L. *The Grand Design*. New York, Random House Publishing Group, 2010, p. 52 ff.
[176] Keil, G. *Willensfreiheit und Determinismus.* 2nd ed. Ditzingen, Philipp Reclam jun. GmbH & Co. KG, 2009, p. 23 ff.
[177] Fischer, E/Vossenkuhl, W. *Die Fragen der Philosophie: Eine Einführung in Disziplinen und Epochen.* Munich, C.H. Beck, 2003, p. 27.

opinion is highly important, as it states that sanity is intellectual potency which is not reliant on a potential for perception, delivered by senses. Even though modern science is able to explain how brains function, there is still no explanation for the major factors leading to self-awareness, and how brains create meaning out of information. Criminal law for instance, would be heavily questioned if all capabilities of human beings in terms of deciding freely and correctly between lawful and unlawful were doubted.[178] In the German criminal legal science, several opinions on human being's freedom of will exist, however the prevailing opinion in the literature holds that the basis of our legal and social system is the self-responsibility of human beings, because science may not prove the freedom of our will but may not refute it either. Human beings are self-responsible for the order of their existential organisation, as a compulsory meaning of life.[179] The similarity of the discussions in both of the scientific fields presented lies in the question of whether all human declarations of will and actions (similar to the decision-making and problem-solving abilities of autonomous intelligent agents) are based exclusively on deterministic physical and biological laws, or whether, despite biology and physics, there is still free will and free agency of humans. How exactly the correlation between the cognitive abilities and the characteristics of human beings developed by environment, education and experience, and how these characteristics are influenceable is not yet scientifically clear. In contrast to the holistic philosophical consideration, law is rather linked to the appearance of deeds, or rather to the behaviour patterns towards other parties.[180]

It is evident that autonomously acting software agents are capable of solving complex problems which need a certain degree of rationality based on their intelligence, which exhibit problem solving capabilities close to that of humans. However, it needs to be considered that the current state of development allows AI software agents to perform certain subtasks autonomously, while a human being remains responsible for his actions. All of these independent subtasks are

[178] Küng, H. *Anfang aller Dinge.* Munich, Piper Verlag, 2008, p. 197.
[179] Lackner, K./Kühl, K. In. Lackner, K./Kühl, K. *Strafgesetzbuch.* 28th ed. Munich, C.H. Beck, 2014, Vor. § 13, recitals 24, 26.
[180] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 144.

embedded into an overall organisation and, thus, are dependent upon that specific organisation and control of human beings.[181] Here, the software agents must be capable of perceiving certain situations occurring from changes in the environment, having a positive or negative influence on their goals, and involve these changes in the considerations of actions, granted by the software.[182] The behaviour of software agents is based on mathematical formulas or algorithms. However, their capacity of action is designed in a primarily autonomous and flexible manner that may be altered by their functionality and behaviour, allowing them to adapt to changes in their surroundings. Such capabilities require the development of appropriate mathematical formulas and programming.[183] This means that software is some kind of a prescription, or algorithm consisting of a deterministically programmed sequence of orders, which is formulated in a machine code of the computer, capable of carrying out tasks exactly as intended. When algorithms are faced the same conditions, they will always produce the same outcome. Thus, the question arises whether such technical procedures, or such command sequences, despite their autonomy, have any similarities with the legally relevant decisions of human beings, which are supposed to be based on free, rational and reasonable considerations, relating to the given case. It needs to be kept in mind that the capability of problem solving is based on a flexible algorithm. Human beings do have influence on the design of the flexible algorithm, but abandon a part of their influence by granted autonomy, allowing for cooperatively sharing problem solving tasks with the software agents. The difference between autonomous software-agents and non-autonomously programmed software is that they possess learning and adapting capabilities towards the user's behaviour.[184]

---

[181] Kirn, S. *Integration von Organisation und Informationssystem: Benötigen wir eine Re-Vitalisierung des maschinellen Aufgabenträgers?* Work report, Technische Universität Ilmenau, 1996, pp. 44 ff.

[182] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 144.

[183] See Kirn, S. *Kooperierende intelligente Softwareagenten.* In. *Wirtschaftsinformatik.* 44(1), 2002, 53-63.

[184] Cornelius, K. *Vertragsabschluss durch autonome elektronische Agenten.* In. *Multimedia und Recht*, 5(6), 2002, pp. 353–355; Koch, F. *Internet-Recht.* 2nd ed. Berlin, De Gruyter Oldenbourg, 2005, § 3 III 1b; Heckmann, D. *juris PraxisKommentar Internetrecht.* 1st ed. Saarbrücken, juris, 2007, chapter. 4.1, recital 38 ff.

8.4. Non-Human Entities and Electronic Personhood

Concerning robots, the question of legal responsibility, especially in the sense of criminal law is difficult to answer. In civil and criminal law, people are to be held liable if damage is the result of their own actions. Applying these traditional rules to autonomous systems is more difficult. Modern programming allows AI not only make experiences, but also to analyse them. It further allows their users have an effect on them and to even make autonomous choices, on the basis of the actual circumstances. This results in the increasingly less foreseeable behaviour of autonomous systems, which makes it impossible to link this back to a specific action of a human.[185]

It can be argued that it no longer makes sense to differentiate between humans and machines. Humans and machines should be considered simultaneously instead, because technology has indissolubly connected them together. However, technology has not yet reached this far. Either way, this does not mean, that human beings and robots should be accorded the same rights and responsibilities, but rather to make the fact understandable that human beings (especially users) may be at least partially not liable when using some type of robots.[186]

It generally makes sense to differentiate between software agents and physical robotic machines, especially when focusing on non-human agents. The legal status of software agents is subject to intensive debate. It should be possible to apply some ideas of the legal status of software agents to embodied robots, for instance, autonomous systems concluding contracts.[187]

8.5. Electronic Personhood

At this point, the question arises whether legal personhood of AI can provide an efficient legal basis to tackle the problems of responsibility. Usually, legal systems are, to some extent, flexible and may allow for creating and adding a new type of entity to the existing system.[188] Also, the European Parliament has

[185] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 3.
[186] Ibid.
[187] Ibid.
[188] Janssens, L. *A Prospect of the Future: How Autonomous Systems May Qualify as Legal*

already begun to deal with questions regarding new concepts of personhood for AI: "whereas, ultimately, the autonomy of robots raises the question of their nature in the light of the existing legal categories or whether a new category should be created, with its own specific features and implications."[189] Regarding autonomous systems, granting them legal personhood is highly complicated because it brings up many challenging questions, such as: who is authorised to determine which systems may be granted legal personhood? What are the prerequisites to do this?[190]

Looking into the more distant future, it might be a considerable step to give robots or software agents the electronic personhood. As technology advances, there is a significant chance that AI will require legal categorisation of such entities. Nowadays, any traditional concept of responsibility fails as any AI lacks consciousness, which means they continue to be defined as a 'thing'.[191] Today's technology is not sufficiently advanced that according them with the same legal status as humans could be regarded as a necessary measure. Hence, machines should not yet be given the same legal status, though it is possible to establish a specific legal category for machines.[192] The concept of legal personhood is amalgamated as the combination of material and financial responsibilities, reflecting the treatment of human beings under the law. Further, the term 'legal person' and what this term covers is heavily dependent on the respective legislation's decision of what, or which group is to be covered by the term, that is to say, which group is to be accorded status as a legal entity. Obviously, robots are neither animals nor human beings. Despite this, they can, to a certain extent, develop a personality, and therefore a specific kind of decision-making. Giving robots legal personhood would constitute an attribution of personhood, similar to what has already been done for companies. A new classification for machines

Persons. In. Janssens, L./Bayamlioglu, E./Baraliuc, I./Hildebrandt, M. *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*. Amsterdam University Press. 2018, p. 4.
[189] 2018, European Commission, Office Journal, (C252/25), p. 242.
[190] Janssens, L. *A Prospect of the Future: How Autonomous Systems May Qualify as Legal Persons.* In. Janssens, L./Bayamlioglu, E./Baraliuc, I./Hildebrandt, M. *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*. Amsterdam University Press. 2018, p. 20.
[191] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 5.
[192] Ibid.

is necessary primarily because legal personhood would allow for the bundling of all legal responsibilities of the various parties involved, including users, sellers, producers, and others. In this case autonomous machines could be classified as 'electronic persons' and possess specific rights and obligations. However, this could only apply to machines equipped with a certain degree of legal autonomy in certain specific contexts.[193] This concept would especially make sense for AI, strong enough to make decisions on its own, or to in some way be able to interact with people, for instance by closing contracts or even infringing a person's protected rights. Machines would then be recorded in a public register, probably similar to the commercial register, and would obtain their legal status from the point of their registration.[194] This registration process could be similar to the process that natural persons would need to go through, for example after moving into another city or country.

It is also important to consider the question of which types of machines it makes sense to accord personhood. For those machines made for one very specific purpose, with almost no interaction with other humans (even though they would be able to do so), it would not make much sense to grant them legal personhood. Moreover, even in some cases where robots have a high level of interaction with their environment, it is questionable whether legal personhood is actually necessary. Pertinent examples here might include trucks 'employed' by companies, or perhaps cars in general. These types of machines are highly specialised for one specific task and although they might indeed interact significantly with their surroundings, they will probably never be expected to close contracts. Further, there is no reason for such intelligent tools to have legal person status for cases where an autonomous vehicle does harm to a person's legal interest, since the company that uses such machines will be liable for its actions, as long as factors such as production or programming faults can be excluded. Taking Germany as an example, the employee is liable for damages committed by employees, except in cases of wanton and wilful negligence of course.

[193] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 5.
[194] Ibid.

One type of machine that should be granted electronic personhood could be unattended systems performing specific tasks customised to their special character that cannot be controlled in advance. The critical problem here are the difficulties in using traditional measures created by law to attribute responsibility to a specific subject.[195]

8.6. Software Agents

A possible description of a software agent would be a computer program that is able to take on and complete tasks for its user. In this case, the software agent would have a certain amount of AI on which basis it could execute the respective tasks autonomously. To be able to do so, electronic, or software agents need to possess the characteristics of responsiveness,[196] proactivity,[197] the ability to conclude,[198] and the ability to communicate.[199] Further, the software agent would have the ability to interact with its environment, thus making them capable of acting as representatives of their respective users. According to Article 2:101 of the Principles of European Contract Law (PECL), the conclusion of a contract requires the parties' intent to be legally bound, and to reach a sufficient agreement without any further requirement. The intention of a party to be legally bound by a contract is to be determined from the party's statements or conduct as they were reasonably understood by the other party.[200]At this point the question arises, whether an electronic agent or a robot can have the intention to close a contract. Agreements are declarations of intention. Agreements represent a person's will towards another person to produce a specific legal outcome.[201] Interpreting the principles narrowly, the philosophical or neurological questions

---

[195] Iannì, A./Monterossi, M. W. *Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability.* In. *Griffith Law Review.* 26(4), 2017, p. 578.

[196] Murch, R./Johnson, T. *Agententechnologie: Eine Einführung.* In. *Intelligente Softwareagenten auf Informationssuche im Internet.* 1st ed. Boston, Addison-Wesley, 2000, pp. 29 ff.

[197] Brenner, W. et al. *Intelligente Softwareagenten - Grundlagen und Anwendungen.* Berlin, Springer, 1998, p. 27.

[198] Cornelius, K. *Vertragsabschluss durch autonome elektronische Agenten.* In. *Multimedia und Recht*, 5(6), 2002, p. 353.

[199] Zarnekow, R. *Softwareagenten und elektronische Kaufprozesse*. Berlin, Springer, 1999, p. 22.

[200] Art. 2:102 PECL.

[201] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 3.

concerning consciousness and, thus, the existence of a will of robots might be superfluous, as Article 2:103 PECL sets out that a sufficient agreement is given if (a) the agreement has been sufficiently defined by the parties so that the contract can be enforced, or (b) if the terms can be determined under these Principles. However, taking German law as an example, the declaration of will is the basis for closing a contract. Thus, the subjective minimum prerequisite is the existence of human will to close a contract.[202] It is for this reason that the specific declaration of will of an electronic agent is not yet possible in Germany. Electronic agents have no 'legal life', which excludes them from personhood in German law.[203] This requires software agents to possess the ability of social behaviour, and the ability for the formation of will. However, this requires awareness of the software agent's own existence,[204] which means that the agent not only has to understand the consequences of its contractual actions, but also needs to possess a minimum capacity of judgement,[205] which is not currently fulfilled by any software or robot. However, using software agents or robots to conclude a contract might be legally evaluated as the use of a tool.[206]

Generally, software agents are seen as a tool.[207] Human authors use software as a tool to create a statement. This type of statement is known under the name of 'agent statement', which is considered to be a variety of the 'computer statement'[208] These electronic agents are advanced, intelligent programs, which are

---

[202] Hefermehl, W./Soergel, H-T. *Bürgerliches Gesetzbuch mit Einführungsgesetz und Nebengesetzen (BGB),* vol. 2, 13th ed. Köln, Kohlhammer, 1999, Vor. § 116, recital 20.
[203] Cornelius, K. *Vertragsabschluss durch autonome elektronische Agenten.* In. *Multimedia und Recht*, 5(6), 2002, p. 354.
[204] Ibid.
[205] Brox, H./Walker, W-D. *Allgemeiner Teil des BGB* 37th ed. Munich, Verlag Franz Vahlen, 2013, § 12 recital 259; Schiemann, G. In: Staudinger, J. v. *BGB Eckpfeiler des Zivilrechts – Kommentar zum Bürgerlichen Gesetzbuch und Nebengesetzen.* Berlin ottoschmidt – De Gruyter, 2022, chapter. C, recital D 111; Müller-Hengstenberg, C. D./Kirn, S. *Intelligente (Software-) Agenten: Eine neue Herausforderung unseres Rechtssystems Rechtliche Konsequenzen der "Verselbstständigung" technischer Systeme.* In. *Multimedia und Recht.* 17(5), 2014, p. 308.
[206] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 3; Gitter, R. *Softwareagenten im elektronischen Geschäftsverkehr - Rechtliche Vorgaben und Gestaltungsvorschläge.* Baden-Baden, Nomos, 2007, p. 172; see also Wettig, S. *Vertragsschluss mittels elektronischer Agenten.* In. *Schriften zur Rechtswissenschaft*. vol. 142, Berlin, Wissenschaftlicher Verlag Berlin, 2010.
[207] Gitter, R. *Softwareagenten im elektronischen Geschäftsverkehr - Rechtliche Vorgaben und Gestaltungsvorschläge.* Baden-Baden, Nomos, 2007, p. 172.
[208] Spindler, G./Schuster, F. *Recht der elektronischen Medien*. 4th ed. Munich, C.H. Beck. 2019, Vor. §§ 116 ff., recital 10; Günther J-P. et al., *Issues of privacy and electronic personhood in*

able to perform tasks with a high level of autonomy, while being able to react to the given circumstances of their environment.[209] Despite their, by now, high complexity and autonomy, electronic agents are still not recognised to have legal capacity.[210] Electronic, or software agents have a wide range of possible applications, even though human beings continue to be behind their development, determining the core of their autonomous decision-making processes, meaning humans remain accountable.[211] This means that, concerning the state-of-the-art technology, computer statements are retraceable to their respective users, as the user gives general recognition and consent in the use of a computer statement.[212] However, the question whether the statement of a software agent may be categorised as a computer statement has been subject to debate. It is a matter of fact that software agents are intelligent, autonomous and have the ability of interaction, which enables them to affect the circumstances crucial for the conclusion of the contract. Therefore, the outcome of a software agent's action is not fully foreseeable, even if the user knows how the AI is programmed and every further pertinent circumstance. This constitutes a significant difference to computer statements, as the actions and outcomes of such statements are always foreseeable. Further, there is no possibility for a user to influence the actions of the software agent, which is a required characteristic of a computer statement.[213]

8.7. Conclusion: Electronic Personhood as the Solution?

The introduction of the electronic personhood seems to be a competent solution for the diffusion of responsibility. Legal personhood, as we know it from companies means the bundling of legal capacities, financial and material

---

robotics. In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 3.

[209] Brenner, W. et al. *Intelligente Softwareagenten - Grundlagen und Anwendungen.* Berlin, Springer, 1998, p. 23; Gramlich, L. et al. *Rechtshandbuch B2B Plattformen – Rahmenbedingungen elektronischer Marktplätze*. Munich, C.H. Beck, 2003, § 1, recital 48.

[210] Müller-Hengstenberg, C. D./Kirn, S. *Intelligente (Software-) Agenten: Eine neue Herausforderung unseres Rechtssystems Rechtliche Konsequenzen der "Verselbstständigung" technischer Systeme.* In. *Multimedia und Recht.* 17(5), 2014, pp. 307 ff.

[211] Spindler, G./Schuster, F. *Recht der elektronischen Medien*. 4th ed. Munich, C.H. Beck. 2019, Vor. §§ 116 ff., recital 10.

[212] Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012. (2012), p. 4.

[213] Ibid.; see also John, R. *Haftung für künstliche Intelligenz – Rechtliche Beurteilung des Einsatzes intelligenter Softwareagenten im E-Commerce.* In. *Recht der Neuen Medien.* vol. 45, Hamburg, Dr. Kovac. 2007.

responsibilities. Legal persons are treated as humans under the law, although, in other respects are not given the legal status as human beings receive. The category of legal persons does not cover all groups, as it is a decision made by law, to which group a legal status shall be given.[214] Corporate liability is proven to be a sufficient solution to hold at least the person doing wrong as liable, while nobody is made responsible for any damage resulting from the company's actions. As mentioned above, many countries have even established criminal liability for corporations.[215]

A similar approach for machines which are autonomous to some extent is thinkable. Theoretically, robots can develop an artificial personality. They further can develop a certain scope of action and decision-making.[216] This shows that it is possible to create a legal status as a tangible symbol for the cooperation of everyone creating and using that robot.[217] From an internal standpoint, creating new legal entities with distinct legal obligations is not a significant issue. This would require these particular autonomous machines to have a certain degree of legal autonomy. According legal personhood to machines would be highly beneficial, as it would bundle together all legal responsibilities of the various parties. This approach would affect civil law, since judgments could be handed down against electronic persons directly. The specific judgments would be covered by the electronic persons assets, paid in by the parties involved in the creation and training process. It is conceivable that in cases where a machine's malfunction is caused by a severe lack of care or intentional wrongdoing, the transfer of payment could be directed to one of the parties responsible for the machine's development or use.[218]

However, the legal concept of the electronic personhood is not based on robots characterised as artificial humans.[219] To establish such a concept, it would

[214] Beck, S. *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 141.
[215] Ibid.
[216] Hanisch, J. *Haftung für Automaten.* In. *Internationale Göttinger Reihe.* Vol. 19, Göttingen, J.-P. Cuvillier, 2010, p. 208.
[217] Beck, S. *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 141.
[218] Ibid., p. 142.
[219] Ibid.

need to be grounded in ontological factors that compare it to humans, evaluating traits like mobility, sensory perception, capacity for learning, intentionality, identity, capacity for reasoning, responsiveness to reasoning, ability to hold second-order desires, mental soundness, and other related attributes.[220] The difficulty here is that even human beings have still not been shown to fulfil the characteristics of consciousness.[221]

While corporate criminal liability is applied in many countries, it is opposed in others, such as Germany. The academic discussion is divided between two arguments: one strand argues that such a legal concept is based on an inner logic of the legal system, while the other maintains that there are insufficient similarities between corporations and humans, meaning that criminal law is not applicable.[222] This latter point is especially important considering the fact that there is a significant difference between corporations and robots. Robots directly interact with humans in the real world, and are possibly even able to show emotions and to react in an empathic manner. As a result, these robots may be more easily regarded as social actors, or actual counterparts, instead of fictional entities.[223]

However, the above-mentioned considerations do not imply that the introduction of criminal liability for robots is wrong. Concerning the state-of- the-art in AI, the non-similarity approach is more convincing for now. Sticking with Germany as an example, domestic criminal law would therefore need to undergo considerable changes in favour of introducing new entities.[224]

9. Possible Constellations of Criminal Liability

As the possibility of machines culpability has been negated so far, civil law scholars have searched for arguments towards a possible causal liability or strict

[220] Schweighofer, E. et al. *Auf dem Weg zur ePerson: aktuelle Fragestellungen der Rechtsinformatik.* In. Schweighofer, E. et al. *Schriftenreihe Rechtsinformatik*. vol. 3, Vienna, Verlag Österreich, 2001, p. 143.
[221] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 142.
[222] Ibid.
[223] Ibid.
[224] Ibid.

liability of the operator. However, this solution is not competent to eliminate all legal difficulties in application.[225] Based on traditional or common doctrines of criminal law, robots cannot be held criminally liable due to their absence of personhood and free will, indeed, robots cannot even be seen as capable of acting and, thus, cannot be liable. For these reasons, potential criminal responsibility is attributed either to the producer, programmer, or to the operator of the machine.[226] In light of this fact, further analysis of criminal liability for deeds of robots requires the differentiation of liability between the producer, programmer and the operator on the one hand, and the liability of the robot itself on the other.[227]

Concerning this, problems of negligence, omission and perpetration which might be the result in the context of such constellations, is discussed in this chapter. Further, in the next step, the possibility of robot responsibility is reviewed. Here the question arises whether robots or intelligent agents could be criminally liable today, or in the future. Within the academic debate on the criminal liability of robots, there is no avoiding careful examination of the doctrine of criminal liability and questioning the theory of criminal justice.[228] However, discussing this question is highly challenging, since there is no internationally applicable principle or doctrine in criminal responsibility. Most nations follow their own doctrines of criminal law and liability, but ultimately share many similarities. At best, some differences may be pointed out here, sticking mostly to the numerous similarities. While private law, concerning liability, mainly focuses on the compensation of mostly monetary damages, criminal law rather has its focus on the personal accusation, which is to be compensated by punishment. Even though academic opinions and the conceptions concerning the specific purpose of punishment are highly divergent, the question of who is to be held criminally liable is of key interest. Whether society will hold the programmer, producer, operator

[225] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 173; see also Freytag, U. *Sicherheitsrechtliche Aspekte der Robotik.* In. *Sicherheit & Recht/Sécurité & Droit*. 02/2016, pp. 111 ff.
[226] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 173
[227] Ibid., p. 174.
[228] Ibid.

or the robot itself criminally responsible will have major impact on future development in this particular technology.[229]

9.1. Criminal Responsibility of the Programmer, Producer and the Operator

According to the media, autonomous cars promise an increase in safety.[230] The car industry is on the cusp of inventing fully autonomous cars, where the driver does not have to actively watch the actions of his car. In a couple of years, drivers might even be allowed to even sleep while the car drives safely.

In May 2016, this idea was called into question after an accident where an autonomous, possibly speeding car, crashed into a white truck after not braking. The 'driver' of the autonomous car died.

If the driver of the autonomous vehicle has to take responsibility for the cars misconduct, when the control is temporarily transferred to the autopilot, from the driver's perspective, this is a good argument against its use. The possible consequences of such a transfer of control are not legally regulated.[231] *Eric Hilgendorf* has referred to this situation as a "control dilemma,"[232] thus directing attention mainly to the predicament of drivers who seem to be caught between the relief offered by the automotive industry and the obligation imposed by the law.[233] In between, the driver is in charge of being ready to take over the control of the vehicle any time.

Here the question arises, whether a solution needs to be found for the driver's control dilemma. When technology reaches a point where the driver of a car is no longer obliged to assume control of the vehicle at any time, the

---

[229] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 174.

[230] https://www.zeit.de/mobilitaet/2021-11/autonomes-fahren-kuenstliche-intelligenz-verkehrssicherheit-unfall.

[231] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 228.

[232] Hilgendorf, E. *Automatisiertes Fahren und Recht.* In. 53. *Deutscher Verkehrsgerichtstag 2015*, Köln, Luchterland, 2015, pp. 67 ff.

[233] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 227.

liability is handed over to the autopilot. Drivers should be able to respond in a case of an accident. Still, the reason for the result was not their fault but the fault of the autonomous driving system.[234]

However, evidentiary problems remain. It will be difficult for drivers to provide evidence to prove that the autopilot was turned on when the accident occurred. The reason behind this evidential problem is that only the producers of the vehicles have access to the log data which are saved in the car. Drivers are not able to open these files and producers have almost no incentive to make these files open to the users. Further, even if the users were to obtain access to this information, they most probably would not be able to evaluate it.[235]

International law has already provided the pathway for assisted and partially automated driving. The Vienna Convention on Road Traffic requires that the vehicle is controlled by a driver,[236] thus, by a human being. Things are different, when cars reach a higher degree of autonomy, as the degree of automation allows the driver to occupy themselves with other matters while driving.[237] In Germany, this degree of autonomous driving was therefore highly problematic.[238] In 2016, changes to the Vienna Convention on Road Traffic paved the way for the introduction of autonomous vehicles.

Still, many drivers and also politicians welcome future developments in autonomous driving because they see these changes as advancements towards more safety and comfort in road traffic. In this matter, they are also pushing to clarify the question as to whether the drivers, as possible primary beneficiaries of the collaboration between human beings and machines, are always to be liable for any damages that may result from the collaboration.

---

[234] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 229.
[235] Ibid.
[236] Art. 8 paragraph 1, 5; Art. 13 Vienna Convention on Road Traffic.
[237] Deutscher Bundestag, *Autonomes und automatisiertes Fahren auf der Straße – rechtlicher Rahmen.* In. *Wissenschaftliche Dienste*. 3000 111/18, 2018, p. 5.
[238] Lutz, L. S. *Autonomes Fahren als rechtliche Herausforderung*. In. *Neue Juristische Wochenschrift*. 3/2015, pp. 119 ff.

However, even though drivers are theoretically allowed not to drive, they are obliged to be prepared to intervene and to take control of the car at any time the self-driving car requests this.[239]

9.2. Diffusion of Responsibility

Regarding robotics, it is important to make clear that the traditional regime of negligence already reaches its limits when applied to solely one of the potential parties. There are several reasons for this, one being that there is not yet an established social normative framework for robotics as an emerging technology.[240]

However, responsibility of one or more of the parties involved could be challenged even more soon since robots are able to adapt. These abilities could make the robots conduct unpredictable to some extent.[241] A significant question is how legal responsibility changes when robots have autonomous reactions. As the robot continues to gain experience on its own, its behaviour becomes increasingly unpredictable and cannot be fully planned.[242] From a certain extent of a robots autonomy, the question arises whether every mistake made by a robot is necessarily caused by a wrongful deed of one of the 'involved' parties in the legal sense.[243] By allowing robots, adaptive and capable of learning, to interact with humans without supervision, there is a high probability that they may react in unpredictable ways to new inputs. Therefore, if a robot, in such an environment, causes damage due to its reaction, it is not a satisfactory answer that the damage was caused by a wrongful act on behalf of the programmer, producer or the user.[244]

---

[239] Example of German Law: Section 1b (1), (2) No. 1 Road Traffic Act.
[240] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 139.
[241] Ibid., p. 140; see also Günther J-P. et al., *Issues of privacy and electronic personhood in robotics.* In. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.* Paris, 2012.
[242] Boscarato, C. *Who is responsible for a robot's actions?* In. Berg, B./Klaming, L. *Technologies on the stand: Legal and ethical questions in neuroscience and robotics*. Wolfpublisher, 2011, p. 383; Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 140.
[243] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 140.
[244] Ibid.

9.2.1. Tolerated Risks of Highly Automated Driving

A look into the history of road traffic laws shows that the liability for dangerous activities underlies a continuous adjustment since the very introduction of automobiles.

Drivers could claim that activating the autopilot of the vehicle might be an accepted risk and that possible damages are the autonomously acting machine's fault and not that of the driver. In the long term, after extensive transition in road traffic to autonomous driving, the rare injuries within the use of autopilots are likely to be seen as a general risk in life.[245]

9.2.2. Foreseeability

As mentioned above, many parties, such as drivers (obviously), and also producers of highly autonomously driving cars hope that the use of autopilots will reduce the potential for accidents. For example, it is a matter of fact that a computer program, unlike human beings, is permanently focused on its driving task. However, today's technology is not yet far enough advanced to react to complex, or unusual situations as well as human beings can. Thus, even though self-driving cars do not suffer from the same issues that may affect the driver's driving capabilities, such as being tired, stress, wanting to write an email, or anything else that may distract from the driver's focus, certain risks remain.[246] Damage may occur, even if the autopilot functions perfectly well, for instance, in traffic situations where something unforeseeable happens, or if the weather conditions where the proper functioning of the different sensors of the car may not be guaranteed. Under such weather conditions, the sensors may not detect the contrasts and colour spectrums and, thus may not make the correct decisions in the given situation.[247] However, neither the car drivers who use this technology nor the producers bringing this technology on the market actually expect that accidents can be prevented outright by this technology. Instead, they hope

[245] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 233.
[246] Ibid., p. 234.
[247] See Kuri, J. *Tödlicher Tesla-Unfall: Autopilot hielt Lastwagen-Anhänger für hohes Schild.* 2016, https://www.heise.de/newsticker/meldung/Toedlicher-Tesla-Unfall-Autopilot-hielt-Lastwagen-Anhaenger-fuer-hohes-Schild-3253449.htm.

that the number of accidents will be diminished significantly, and probably suspect that the number of accidents might even rise during the phase where autonomous and traditional driving coexist in a certain transition period.[248] Tesla, for example, has claimed that its autonomous vehicles solely drive 'more safely' in general than human drivers, but not entirely faultlessly and, a fortiori, this does not mean that they will never cause fatal accidents.[249] From the perspective of criminal law it means that the factual occurrence of fatal accidents is generally foreseeable.[250]

9.2.3. Liability for Side Effects

Regarding robots, the risks not only include damages or mistakes, but also risks of unwanted side effects. Since every new technology comes with the danger of negative side effects, it is not surprising that robots in particular are a topic of intensive discussions regarding such side effects. Thinking about robots assuming tasks such as baby sitting, nursing elderly people, taking over a human's everyday communication, giving psychological advice or even fight our wars threatens humans' familiar perceptions of the 'social'.[251] The possibility of unwanted side effects does not automatically mean that progress should be limited or prohibited. However, when considering the field of robotics, it's important to recognize the accountability for any potential side effects.[252]

9.2.4. Breach of Duty of Care

Applying the general definition of duty of care of drivers to highly autonomous cars pose various difficulties.[253] These problems arise from the fact that a (criminal) duty of care must be formulated in a way that the duty is clear and attainable for the recipient of the duty. The first problem already arises at that point, because the purpose of an autopilot is to autonomously collect, and

---

[248] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 234.

[249] Shahan, Z. *1st Tesla Autopilot Fatality… After 130 Million Miles (Updates).* 2016, https://cleantechnica.com/2016/06/30/1st-tesla-autopilot-fatality-130-million-miles/.

[250] Wohlers, W. *Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen.* In. *Basler juristische Mitteilungen.* 3/2016, p. 117.

[251] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 140.

[252] Ibid.

[253] Wohlers, W. *Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen.* In. *Basler juristische Mitteilungen.* 3/2016, p. 117.

process information from its environment to adapt its behaviour to the respective traffic situation. This process does not need to be comprehensible for the driver, as he is most likely resting in the moment. However, all current autopilots still allow for the possibility of overriding the car and giving the control back to the driver whenever human intervention is, or seems to be, necessary.[254] Nevertheless, a duty of care would appear illegitimate for a car driver to always monitor and potentially correct the autopilot, given the option of automated driving. The driver of an autonomous vehicle cannot know with certainty whether the car will interpret a suddenly appearing obstacle or a dirty road sign correctly or incorrectly while he switches to autopilot mode. The driver will know this earliest when he himself recognizes the misjudgement and any resulting dangers, either because he is recalled by the autopilot or because he realizes that he should have been recalled by it. This lack of control in real time is one downside of the relief which autonomous agents should be bringing to humans. A duty of care in such a way that a car driver must always correct the system's errors before damage occurs is impossible for a human to fulfil.[255] A duty of care would make more sense in cases where the driver can turn on the autopilot when he has no inducement to doubt the functionality of the autopilot, in situations where he has to be ready to take over control, to react to warning signs or to react immediately etc.[256] However, many sensitive points remain in the link between human beings and machines. For example, how liability is to be attributed between the driver, owner, producer and even data providers is still not clarified. Possible useful solutions depend on technical development, and the purpose, for instance, for individuals or for car sharing models.[257]

9.2.5. Contributory Negligence Liability

Next to the duty of care, it should be clarified whether turning the vehicle to autopilot might lead to contributory negligence liability per se, even though the

---

[254] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 235.
[255] Ibid., p. 236.
[256] Zurkinden, N. *Strafrecht und selbstfahrende Autos - ein Beitrag zum erlaubten Risiko.* In. *recht - Zeitschrift für juristische Weiterbildung,* 03/2016, pp. 4 ff.
[257] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 236.

driver knew, or should have known that he may not have control over the causal course where the factual event occurs. The concept of contributory negligence liability, in German law, was developed for cases where taking over dangerous activities which require special knowledge, skills, or special organisation, in order to keep the risk for dangers to occur as low as possible.[258] This kind of liability is not meant to hand over a de facto impossible control to a person. Contributory negligence liability allows us to find an adequate solution much rather if somebody acts contrary to his duty, because he takes over tasks to which he is not suited, whilst being doable per se. Examples of such tasks include performing a specific medical treatment for which the medical doctor lacks familiarity or routine.[259] However, the concept of contributory negligence liability cannot bypass the lack of control, which comes along with autonomous driving, in which the driver realises that the autopilot cannot react adequately in a certain situation. Here the driver is not accused of taking on a risky task for which he is not suited. The point is the authorisation of a technology with the key feature of being uncontrollable by human beings to correct its actions.[260]

9.2.6. Autonomous Driving and Acceptable Risk

The purpose of criminal liability for negligence, however, is not to prohibit any possibly risky activity.[261] In many areas, risky methods and proceedings are highly appreciated. Other areas may at least accept such behaviour.[262] Road traffic is an extremely good example for the acceptance of acceptable risk, due to the overriding public interest. This societal decision, however, is by no means self-evident, looking at the fact that thousands of people being injured and killed

[258] Kudlich, H. In. Heintschel-Heinegg, B. v. *Beck'scher Online Kommentar StGB*. 32nd ed. Munich, C.H. Beck, 2016, section 15 recital. 66.
[259] Seelmann/Geth, K. *Strafrecht Allgemeiner Teil.* 6th ed. Basel, Genf, München, Helbing & Lichtenhahn, 2016, pp. 170 ff.
[260] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 237.
[261] Frisch, W. *Vorsatz und Risiko - Grundfragen des Tatbestandsmäßigen Verhaltens und des Vorsatzes: Zugleich ein Beitrag zur Behandlung außertatbestandlicher Möglichkeitsvorstellungen*. Köln, Berlin, Bonn, München, Heymann, 1983, p. 59; Prittwitz, C. (1993), p. 371 ff.
[262] Prittwitz, C. *Strafrecht und Risiko - Untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft.* Vol. 22, 1st ed. Frankfurt am Main, Vittorio Klostermann, 1993, pp. 307 ff.

in road traffic every year.[263] Nevertheless, a liability exemption within the scope of permitted risk is generally accepted here. Assessing autonomous driving by comparing the pros and cons, drivers switching their cars into autopilot mode could argue that any, hereby probably newly established risks might be abolished by the desired use of this new technology, which justifies the use of it. However, such ideas might only work out at the point where society widely accepted the fact that autonomous cars drive at least as safely as human beings. This requires that cars not only master routine situations better than human beings, but also exceptional situations.[264] Social evaluations of traffic safety are time- and culture-dependent and are constantly in a process of development.[265] Evidence for this are the highly variable differences in speed limits all across countries.[266] In future there is a good chance that drivers might be obliged to switch their vehicles to autopilot mode if autonomous driving is proven to be safer.[267] It might even come to the point that cars will not even provide human beings the possibility to handle the vehicle anymore.

The deciding factor for risk distribution between driver, producer and other parties in autonomous driving such as data, and software providers, will be the way duties are arranged, especially between the car producers, software developers and other producers. Society will probably only be open to accept residual risks if these are minimised by handling such technology with the necessary care on the part of the users, and if the technology passes through proper tests and continuous monitoring before and after release to the markets.[268]

---

[263] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 240; In 2023, in Germany alone 2776 people died in road accidents and 358,665 were injured in a total number of 2,403,366 accidents so far. Source: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/_inhalt.html. [Last Access: 20.04.2023].

[264] Ibid.

[265] Ibid., p. 241.

[266] Prittwitz, C. *Strafrecht und Risiko - Untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft.* Vol. 22, 1st ed. Frankfurt am Main, Vittorio Klostermann, 1993, pp. 291 ff.

[267] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 241.

[268] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft.* 126(3), 2014, p. 584.

9.2.7. Ordinary Life Risk

The last resort for the driver to avoid criminal liability might be the reference to ordinary life risk. If every car in the future had an autopilot as standard equipment, and this was used by most drivers, they would be generally known and accepted together with their benefits and risks. Most probably the functionality of the autopilots will increase by machine learning exponentially, thanks to the constantly growing amounts of data. Still, in rare cases, the autopilots might make mistakes, for example, if the weather conditions are bad. There might be a point where society accepts that, under such conditions, it is dangerous to activate the autopilot, which is 'confused' by the inadequate visibility, just as society has learned that it is dangerous to walk over the street without taking a certain amount of care.[269]

9.3. Blaming the AI as Defence?

Due to the difficulties in finding proper measurement for duties of care and provisions for the limits of accepted risks, shifting to one specific and new defence comes into consideration. The driver might refer to the autopilot taking control over the car. Whenever the driver switches to autopilot mode, as he is allowed to do so, the driver may be seen as some type of autonomously acting third party. The damaging actions of this may not be attributed to the driver in this case, thus the results (injury, or death) would not be the driver's fault.[270] This approach may seem to be convincing for drivers to actually use the autopilot, however there are still obstacles to overcome: even if society and law accepts the idea of seeing autopilots as 'co-pilots', the human driver is not automatically safe from criminal prosecution. Criminal law knows examples, where people can be held criminally liable for the deeds of someone else.[271]

---

[269] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, pp. 578 ff.

[270] Gless, S. *"Mein Auto fuhr zu schnell, nicht ich!" - Strafrechtliche Verantwortung für hochautomatisiertes Fahren.* In. Gless, S./Seelmann, K, *Intelligente Agenten und das Recht.* 1st ed., vol. 9, Baden-Baden, Nomos, 2016, p. 243.

[271] Ibid.; Wohlers, W. *Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen.* In. *Basler juristische Mitteilungen.* 3/2016, pp. 122 ff.

It may seem strange to redirect the target of criminal prosecution by referring to a mistake of the autopilot, even if it may not be punished any way.[272] As mentioned above, criminal law for robots does not yet exist. However, the claim that not the driver's criminal liability, but the autonomous action of the autopilot is the critical point is mostly invalid. Interrupting the chain according to which liability shall be attributed if an autopilot drives a car leads to certain doubts. Due to the autopilots lack of own will, we do not accredit them the ability of purposeful acting. Thus, its actions appear pre-programmed and not as something stemming from its own will.[273] With an autopilot, we cannot lead a moral discussion about its decisions and behaviour.[274]

Modern penology (or criminal law doctrine) lacks convincing criteria to distinguish the targeted actions of human beings from pre-programmed executions by robots.[275] Taking an autopilot as an example, we see that it is able to drive the car successfully, maybe even better than human beings. It is able to manoeuvre the car to its destination safely, by following traffic rules, performing intermediary steps such as avoiding collisions, speeding up, and avoiding crashes with other cars or obstacles. To achieve this, the autopilot continuously has to make decisions between a varying number of different options, for example, braking, overtaking and changing lanes. However, after the manifold choices the autopilot makes, none of these choices seem 'natural', but rather like pre-programmed decisions between several options. Such a natural decision might be accepted as natural, if the autopilot wilfully sets its own goals and makes its own decisions accordingly. Additionally, it would have to be aware of the fact that its decisions may have a strong social impact. However, as this type of autopilot does not yet exist, the intelligence it uses is a rather simple form of AI, which is not strong enough to support the above-mentioned feature. The weakness of that type of AI leads us to another weakness of an anthropocentric dogma: should the redirection

[272] Gless, S./Weigend, T. *Intelligente Agenten und das Strafrecht.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft*. 126(3), 2014, p. 577.
[273] Stratenwerth, G. *Schweizerisches Strafrecht, Allgemeiner Teil I: Die Straftat. 4th ed*. Bern, Stämpfli Verlag AG, 2011, pp. 128 ff.
[274] Hörnle, T. *Guilt and Choice in Criminal Law Theory.* In. *Bergen Journal of Criminal Law and Criminal Justice,* 4(1), 2016, p. 17.
[275] For the basics, see: Searle, J.R. *Minds, brains, and programs.* In. *Behavioral and Brain Sciences.* 3(3), 1980, p. 417 ff.

of criminal liability only be possible if a cognitive, aware, and perhaps even empathic person or agent assumes control of the vehicle? Yet even if the autopilot acts wilfully, the human 'driver' is not necessarily safe from criminal liability. In German criminal law this is not possible due to the so-called 'Regressverbot'. This means that the perpetrator cannot pass his liability to another, third person, if the third person steps in and influences the course of events initiated by the perpetrator. Thus, deliberate intervention of a third person, into the cause of events, may not free the first person, who caused the course of events, from criminal liability.[276] The prevailing opinion in criminal law literature denies such a possibility.[277] Instead, it is possible to hold both parties criminally liable, independently from each other, if they are objectively both at blame for the event.[278]

9.4. Negligence

9.4.1. Introduction of a Real-Life Example

Scenarios of AI systems injuring or killing human beings are nothing new. A particular case of note took place on 18 March 2018, where Uber tested an autonomous car based on a modified Volvo operating with a self-driving system. The test led to an accident with a pedestrian in Arizona. A 49-year-old woman was hit by the autonomous car and did not survive. The test vehicle was occupied by one operator.

The description of the case is as follows:

> Uber had equipped the test vehicle with a developmental self-driving system. The system consisted of forward- and side-facing cameras, radars, LIDAR, navigation sensors, and a computing and data storage unit integrated into the vehicle. Uber had also equipped the vehicle with an aftermarket camera system that was mounted in the windshield and rear window and that provided additional front and rear videos, along with an inward-facing view of the vehicle operator. In total, 10 camera views were recorded over the course of the entire trip. The self-driving system relies on an underlying map that establishes speed limits and permissible

[276] Puppe I, In: Kindhäuser, U. et al. *Strafgesetzbuch.* 4th ed. vol. 1, Baden-Baden, Nomos, 2013, StGB, Vor. § 13, recital. 167 ff., with further extensive references.
[277] Eisele, J. In: Schönke, A./Schröder, H. *Strafgesetzbuch.* 29th ed. Munich, C.H. Beck, 2014, Vor. § 13ff. recitals 77, 100 ff.
[278] Berster, L. *"Amoklauf von Winnenden" - zur Fahrlässigkeitshaftung neben der volldeliktischen Vorsatztat eines anderen.* In. *Zeitschrift für Internationale Strafrechtsdogmatik.* 12/2012, p. 623.

lanes of travel. The system has two distinct control modes: computer control and manual control. The operator can engage computer control by first enabling, then engaging the system in a sequence similar to activating cruise control. The operator can transition from computer control to manual control by providing input to the steering wheel, brake pedal, accelerator pedal, a disengage button, or a disable button. The vehicle was factory equipped with several advanced driver assistance functions by Volvo Cars, the original manufacturer. The systems included a collision avoidance function with automatic emergency braking, known as City Safety, as well as functions for detecting driver alertness and road sign information. All these Volvo functions are disabled when the test vehicle is operated in computer control but are operational when the vehicle is operated in manual control. According to Uber, the developmental self-driving system relies on an attentive operator to intervene if the system fails to perform appropriately during testing. In addition, the operator is responsible for monitoring diagnostic messages that appear on an interface in the centre stack of the vehicle dash and tagging events of interest for subsequent review. On the night of the crash, the operator departed Uber's garage with the vehicle at 9:14 p.m. to run an established test route. At the time of the crash, the vehicle was traveling on its second loop of the test route and had been in computer control since 9:39 p.m. (i.e., for the preceding 19 minutes). According to data obtained from the self-driving system, the system first registered radar and LIDAR observations of the pedestrian about 6 seconds before impact, when the vehicle was traveling at 43 mph. As the vehicle and pedestrian paths converged, the self-driving system software classified the pedestrian as an unknown object, as a vehicle, and then as a bicycle with varying expectations of future travel path. At 1.3 seconds before impact, the self-driving system determined that an emergency braking maneuver was needed to mitigate a collision (see figure 2). 2 According to Uber, emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior. The vehicle operator is relied on to intervene and act. The system is not designed to alert the operator. The self-driving system data showed that the vehicle operator intervened less than a second before impact by engaging the steering wheel. The vehicle speed at impact was 39 mph. The operator began braking less than a second after the impact. The data also showed that all aspects of the self-driving system were operating normally at the time of the crash, and that there were no faults or diagnostic messages. Several Uber self-driving system cameras captured the crash event. The videos were reviewed by the NTSB and the parties to the investigation. The forward-facing videos show the pedestrian coming into view and proceeding into the path of the vehicle. The videos also show that the pedestrian, once visible, did not look in the direction of the vehicle until just before

> impact. The videos show that the pedestrian was dressed in dark clothing and that the bicycle did not have any side reflectors. The bicycle had front and rear reflectors and a forward headlamp, but all were facing in directions perpendicular to the path of the oncoming vehicle. They further show that the pedestrian crossed in a section of roadway not directly illuminated by the roadway lighting. The inward-facing video shows the vehicle operator glancing down toward the center of the vehicle several times before the crash. In a post-crash interview with NTSB investigators, the vehicle operator stated that she had been monitoring the self-driving system interface. The operator further stated that although her personal and business phones were in the vehicle, neither was in use until after the crash, when she called 911.[279]

Complex cases like this raise the question of what meaningful control is. Does it mean people's control over machines and systems? Or does it mean that autonomous AI systems limited autonomy, acting within a narrow margin and framework for assessment in which they can evaluate, judge and act in specific situations?[280] Various parties were involved in this case. Thus, the question of who, or what was in control in the exact moment concerns different actors. Is Volvo to blame for delivering a malfunctioning emergency brake system? Is Uber to blame for purchasing the software? Is the operator guilty who monitored the test-vehicle by looking at the screen instead of the road, at the moment the accident occurred? Or is it the autonomous vehicle itself?[281]

These questions concern the emerging problems of liability and bring up the question of who is responsible for the basis of how the specific software works. Further, who can be responsible for the meaningful control over the software and the whole system? A huge number of factors may cause errors. However, errors or failures are mainly caused by the vast amount of possible traffic situations, where many of them are highly complex and even challenging for human beings. These systems must be reliably assessed by the software. There are many examples of autonomous vehicles making the wrong decisions in specific situations.

---

[279] National Transportation Safety Board, Preliminary Report, p. 2, available at: https://www.wsj.com/public/resources/documents/NTSBuber.pdf (last access: 09.01.2023).
[280] Janssens, L. *A Prospect of the Future: How Autonomous Systems May Qualify as Legal Persons.* In. Janssens, L./Bayamlioglu, E./Baraliuc, I./Hildebrandt, M. *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*. Amsterdam University Press. 2018, p. 4.
[281] Ibid.

For example: GM's prototypes of autonomous cars attempted to change lanes to the opposite side of the road. A car equipped with the technology of Google's Waymo was involved in an accident, where it tried to drive the car into streets that were too narrow, and a prototype of Telenav confused a roundabout with a statutory vehicle. Autonomous systems will increasingly make their own assessment frameworks, the more autonomously these systems work. Thus, it will become more complicated to address the question of responsibility.[282] Legislation usually starts from the point that robots are, in terms of civil law, to be qualified as items or things and, thus, are legally not capable of committing a tort. Criminal law, except corporate criminal law, only penalises natural persons. Objects or animals for example cannot be subjects to criminal law. This has to do with the fact that criminal law only declares actions or omissions punishable which can be influenced by will, and such a will-controlled behaviour cannot be attributed to robots due to the current state of technology.[283] The traditional definition of guilt further requires that the person being accused of a criminal action must have had the possibility to behave differently, thus to not have committed the specific criminal act.[284]

Due to the programming and their limited autonomy within their decision-making, robots cannot be accused of acting differently, that is, acting either lawfully or unlawfully according to the current doctrines of criminal responsibility. Hence, currently, robots cannot be held criminally liable.[285] However, from the point where such autonomous machines exist, the criminal liability of, in this case, strong AI is another question.

If a robot performed actions leading to a result relevant under criminal law, and is thus considered to be a criminal offence, the question arises as to who is responsible for the robot's action.[286] Staying with the examples of traffic

[282] Janssens, L. *A Prospect of the Future: How Autonomous Systems May Qualify as Legal Persons.* In. Janssens, L./Bayamlioglu, E./Baraliuc, I./Hildebrandt, M. *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*. Amsterdam University Press. 2018, p. 4.
[283] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 174. According to Swiss, and also German criminal law.
[284] Ibid.; see also BGH, 18.03.1952, GSSt 2/51, recital 17, NJW 1952, 593.
[285] Ibid.
[286] Ibid.

accidents involving autonomous cars, causing property damages or personal injuries as a result of their actions, the specific behaviour may lie in deficient programming. Then, possible negligence of the programmer or producer needs to be examined. If the robot was deliberately programmed that the criminal deed or result is fulfilled, the respective deed could be subsumed as an intentional criminal offence, which can be committed through an active behaviour or omission by the programmer, producer or the user.[287] If the robot is programmed correctly but used improperly, then the criminal responsibility of the operator would be subject to review, either in the sense of recklessness or in the sense of wilful endangering.[288]

At first the question arises whether negligence is a thinkable scenario regarding AI, and how such scenarios might look. Here it shall be assumed that none of the above-mentioned parties had the intention of violating a human being. Thus, a robot violating a human being could lead to criminal liability based on negligence. The aforementioned form of criminal responsibility can be linked to all stages of production, usage, and even research and development.[289]

Taking German criminal law as an example, the prevailing concept implies that the criterion of negligence requires a behaviour which violates the duty of reasonable care.[290] That being said, not every action or refraining from a certain action is enough to fulfil the requirement of negligence, rather it needs a violation of the required due diligence,[291] the creation of a risk of the violation of a legally protected right,[292] or the breach of the permissible risk.[293] Normally, the

[287] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 174.
[288] Ibid., p. 175.
[289] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 138.
[290] Sternberg-Lieben, D./Schuster, F. In. Schönke, A./Schröder, H. *Strafgesetzbuch.* 30th ed. Munich, C.H.Beck, 2019, section 15 StGB, recital 116.
[291] Welzel, *Fahrlässigkeit und Verkehrsdelikte: Zur Dogmatik der fahrlässigen Delikte.* vol. 49, Heidelberg, C.F. Müller, 1961, p. 132.
[292] Ibid., p. 129.
[293] See e.g. Duttge, G. In. Joecks, W./Miebach, K, *Münchener Kommentar zum StGB*. vol. 4, 4th ed. Munich, C.H.Beck, 2020, section 15, StGB, recital. 105-107; Kühl, K. In. Lackner, K/Kühl, K. *Strafgesetzbuch.* 29th ed. Munich, C.H. Beck, 2018, section 15, StGB, recital. 39.

standard of care is specified by the expected situational behaviour of a person.[294] However, there are two major factors in research and production of AI to be considered while determining the standard of care: first, nowadays, only a few standards exist for the relevant areas of robotics. That is to say, it is still difficult to determine who should be given the competence to determine the expected form of behaviour. The slow progress in establishing standards can be attributed to the challenge of determining the expected behaviour for machines that are still undergoing development, given the limited knowledge about the potential risks (both in terms of the risk's nature and severity) that these standards would apply to. Standard-setting organizations are faced with the dual challenge of not only identifying ways to mitigate risks that are deemed inadequate but also determining which risks should be classified as such. In these instances, the benchmark of rationality is often applied to evaluate how a sensible individual would have acted to prevent harm in a similar situation. However, this evaluation does not provide satisfactory answers in complex field such as robotics.[295] Second: non-legal norms are only indicators for whether actions of a person are consistent with the legal standard of care. Further, they are generally developed in accordance with civil liability and not criminal law. Overall social morality also has to be considered. In cases where internal rules align with social expectations and rational standards, and where any involved parties have acknowledged any shortcomings, criminal liability must account for negligence.[296] Besides minimising risks and preventing danger, criminal law further serves the purpose of stabilising the normative consciousness of society regarding deeds that are considered as socially inadequate. In consequence, the danger of a certain action is not sufficient to penalise it. The certain action further needs to violate social-moral rules.[297]

9.4.2. Foreseeability

Another prerequisite for criminal negligence is the question of whether the damage was foreseeable. An offense is established if the damage incurred could

[294] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 139.
[295] Ibid.
[296] Ibid.
[297] Ibid.

not reasonably have been expected based on the general experience of life.[298] The potential dangerousness of machines will rise together with their increasing autonomy. Thus, the fact that they may, later on, cause harm to humans can generally be foreseen during the research phase.[299] Using robots for military purposes or autonomous cars in traffic can be seen as fitting examples. Taking the example of military robots, it is easy to foresee that they will do harm to human beings, since it is their purpose. And within autonomous driving, the violation of human beings seems unavoidable. However, foreseeability is only related to the general possibility of harm. As specific conditions and situations arise, the likelihood of unforeseeable events increases, making it harder to anticipate all potential harms.[300] Due to this, robotics is an opportunity to evaluate how specific the foreseeability has to be. Therefore, the question arises whether it has to be directed towards specific circumstances, causalities or injuries, or whether is it already sufficient to foresee the general possibility of violating humans as such.[301]

9.4.3. Negligence of the Programmer, Producer etc.

9.4.3.1. Violation of the Duty of Care

The question of the criminal responsibility of a person who caused an accident by erroneously programming a robot is already an issue where work processes are supported by highly complex technological tools equipped with a certain degree of automation.[302] As an example, in one factory of the company Volkswagen, a production robot killed an employee by pushing him against a metal plate.[303] In such cases, the fact that nobody tried to 'provoke' the injury,

---

[298] Kühl, K. In. Lackner, K/Kühl, K. *Strafgesetzbuch.* 29th ed. Munich, C.H. Beck, 2018, section 15, StGB, recital. 46.
[299] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 139.
[300] Ibid.; Sternberg-Lieben, D. In. Schönke, A./Schröder, H. *Strafgesetzbuch* 28th ed. Munich, C.H. Beck, 2010, section 15 StGB, recital. 125.
[301] Beck, S. *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood.* In. *Robotics and Autonomous Systems.* 86/2016, p. 139.
[302] Wildhaber, I. *Roboterrecht - Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 2/2017, p. 220; Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 175.
[303] For more details, see Frankfurter Allgemeine Zeitung, (01.07.2015), available at: https://www.faz.net/aktuell/rhein-main/arbeitsunfall-bei-vw-in-baunatal-roboter-toetet-arbeiter-13679358.html.

that is to say, they did not want to cause the offence is indisputable. Cases like this are usually seen as accidents, meaning that only negligence can be considered. However, conditional intent might be necessary to be reviewed if the perpetrator did not wilfully cause the criminal act by his actions but rather that he knowingly accepted the possible outcome. Here, at all events, the question arises whether the programmer might be held liable for negligently programming the robot, leading to a malfunction which ultimately led to the act.[304]

Considering robots, the same rules concerning negligence apply. Negligence must be penalised. However, usually, these rules apply to offences which demonstrate a certain degree of severity. Then, the deciding factor to fulfil prerequisites of an offence through negligence is that the perpetrator must have violated a duty of care through his deeds.[305]

9.4.3.2. Socially Acceptable Behaviour

Taking German, but also Swiss criminal law, as an example, there are actions which are socially adequate and inadequate. According to the ruling of the German Federal Court of Justice, actions that are customary, approved by the general public, and therefore entirely free of suspicion in terms of criminal law within social life, as they lie within the framework of social freedom of action, may not be considered as criminal offense, or at least not unlawful, according to the doctrine of "social adequacy.[306] On the other hand, there are actions which have to be entirely refrained from because they can, due to their hazardous nature or due to their social inadequacy, be considered as criminal act per se. Under this, actions may fall where the perpetrator would not be capable of mastering the associated risks.[307] In such cases the perpetrator would already breach his duty of care by performing the specific action. Consider, for example, the case of a programmer who launches an untested and unpredictable prototype of an autonomous vehicle onto the road as a trial and thereby causes an accident with

[304] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 175.
[305] Ibid.
[306] BGH, 18.02.1970, 3 StR 2/69, recital 11.
[307] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 175.

a pedestrian. Such obviously negligent behaviours are unlikely to pose any problems and are already covered by criminal law today. However, this scenario will probably never occur, as vehicles, other potentially dangerous machines as well as other technological devices will be tested under a controlled and closed environment, such as a laboratory, workshop or, in the case of autonomous cars, closed test tracks, until the risks of this new technology may be reduced to a socially adequate level.[308]

Usually, malfunctions of new technologies should be fixed before these are placed on the market. Still, these new technologies show a higher risk in their first phase after their market launch even if they fulfil the basic standards. The reason for this is that new products are usually not based on a high amount of experience, that is to say they could not be tested for all imaginable scenarios reality might throw at them, which makes them unpredictable to a certain extent. However, this is the 'area' in which most of the difficulties arise. To some extent, risky actions can actually constitute a breach of the duty of care. The prerequisite here is the breach of the maximum permissible, but still within the tolerated (socially adequate) risk. If something goes wrong within this sphere, even though the originator did everything possible to minimise the risk, the action cannot be attributed to him in the sense of a criminal offence.[309] This general rule to prohibit exposing third persons to risks,[310] which is coupled with the foreseeability and avoidability of the specific actions, cannot remain unrestricted in a modern technological society.[311] In some areas of life, a certain degree of exposure cannot be prohibited completely, even though the risk might be predictable and avoidable. This type of exposure constitutes the 'allowed risk'.[312] A good example of this is participating in road traffic, which represents a highly risky activity. Even if the car and driver comply with all traffic laws, the risk of damaging other

[308] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 175.
[309] Ibid., p. 176; Donatsch, A./Tag, B. *Strafrecht I: Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, p. 330 (Swiss criminal law).
[310] Own translation from the German term 'Gefährdungsverbot'.
[311] Sternberg-Lieben, D./Schuster, F. In. Schönke, A./Schröder, H. *Strafgesetzbuch.* 30th ed. Munich, C.H.Beck, 2019, section 15 StGB, recital 144.
[312] Own translation from the German term 'erlaubtes Risiko'.

people's health or property can never be fully eliminated.[313] Because of the usefulness of the risky activity (such as participation in road traffic) to society, the associated risks that cannot be eliminated are accepted by law.[314] Thus, the 'allowed risk' is an expression of the social adequacy, which serves the purpose of avoiding causal actions as far as they are essential to maintain the functioning of society[315] and, in this case, the traffic. In the case of new technologies such as autonomous vehicles, in the context of breaches of duty of care, the question arises on a case-by-case basis whether the use of autonomous vehicles as part of automation processes represents a socially inadequate risk that is not controllable by the manufacturer or programmer, or whether certain risks of this automation must be accepted by society, as long as they do not exceed the limits of acceptable risk. This distinction must also be made by weighing the risks and costs of the technology against the benefits expected for society.[316]

9.4.3.3. Breach of Duty of Care within Socially Adequate Actions

The extent of the allowed risk is determined by the risk potential of the action depending, on the one hand, on legal provisions, for example of the industrial sector's administrative safety regulations and, on the other, if such regulations do not exist, the extent of allowed risks is determined on the general rule to not cause harm.[317] This exact general rule will gain relevance especially in the field of innovative robotics. Such new products inevitably bear new risks and therefore, are a source of danger. Especially the risk of certain malfunctions with potentially fatal consequences cannot be eliminated at the beginning as there is no possibility to fully eliminate such risks. Trying to do so would mean an entire prohibition of innovation. Risks will inevitably remain, unless society desires to completely forego new technology.[318]

---

[313] Sternberg-Lieben, D./Schuster, F. In. Schönke, A./Schröder, H. *Strafgesetzbuch.* 30th ed. Munich, C.H.Beck, 2019, section 15 StGB, recital 144.
[314] Ibid., section 15 StGB, recital 145.
[315] Ibid., recital 146.
[316] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 176, with further references.
[317] Ibid.
[318] Ibid.

9.4.3.4. Attribution and Adequate Causality: The Challenge for Autonomous Robotics and AI

The deeper examination of this topic is based on the example of the programmer to show why criminal liability for a malfunctioning AI product cannot be simply attributed to the producer. The programmer is obliged to do everything in his power to minimise the risk as far as possible, if he does not want to take the risk of being liable for breaching his duty of care.[319]

If a criminal act occurred, which contained a breach of duty of care which the programmer should have considered, the breach must be attributable to the respective programmer. That means that the specific criminal act should have been avoidable by the one specific programmer. Thus, taking the employee who was killed at VW as an example[320], the question arises whether the programmer was able, or should have been able to foresee the threat coming from the malfunction of the robot and then whether he would have been able to prevent the accident by complying with reasonable diligence.[321] To answer this question, Swiss legislation, for example, has formulated a general individually objective standard of care or standard of due diligence which states: "what a conscientious, and considerate person with the education and individual skills of the defendant, in the respective situation did or refused to do."[322]

Finally, it needs to be evaluated whether the breach of the duty of care led to the criminal act. The connection between breach of duty of care and the result of the act is missing if it is clear that due diligence would have been useless, which means that it would not have changed the course of the event.[323] The presence of adequate causality always needs to be evaluated in the context of negligence. According to the Swiss Federal Court, the action of the perpetrator must

---

[319] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 176.
[320] See fn. 298.
[321] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 177.
[322] BGE 122 IV 303 E. 3.a.
[323] Niggli, M./Maeder, S. In.: Niggli, M. A./Wiprächtiger, H. *Basler Kommentar Strafrecht I.* 3rd edition, Basel, Helbing & Lichtenhahn, 2013, Vor. Art. 12 N StGB, recital 117 ff.

be adequate, in accordance with the ordinary course of events and experiences of life, to cause or at least facilitate the respective criminal act.[324] However, it is exactly the criteria of foreseeability and attributability which, especially in context of robotics and AI, may cause difficulties. Concerning programming mistakes, the question arises whether the programmer can ever sufficiently foresee his programming mistakes which lead to a criminal act and thus, whether the programming mistake can be attributed to the programmer. Equally there is the question of whether such a programming mistake, in each case, is in accordance with the ordinary course of events, causative to the criminal act. In context of complex robots equipped with an adaptive system which is capable of learning, adequate causality may barely be existent. However, personal predictability is no longer so easy to assume because the more autonomous a robot can act, the less its "actions" can be attributed to the programmer based on the criterion of predictability.[325]

9.4.3.5. Negligence of the Producer

The criminal liability of the robot's producer for the mistakes of the programmer, or for defects of the product, would also lead to problems in criminal law. Next to the already mentioned difficulties within causality and attribution, some legislations have the possibility to penalise the company[326] if there is no natural person to whom negligence can be attributed.[327] Unlike Swiss criminal law, Anglo Saxon law, for example, has a direct causal attribution of the actions of the staff to the company, or the supervisors respectively, without an independent accusation of organisational fault.[328] If the negligence or fault of the programmer is induced by a defective structure of the company or control measures, the company might be accused of not having taken all precautionary measures to prevent such malfunctions.

---

[324] BGE 130 IV 7.
[325] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 177.
[326] Such as 102 StGB (Swiss).
[327] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 178.
[328] Ibid.

9.4.3.6. Negligence of the Operator

It may be the case that a malfunction or a misconduct of a robot is not caused by the programmer or producer. Misconduct of robots can also be caused by the operator or user of the robot, for instance, by not complying with the rules of usage, or by manipulating the robot in some way. Allowing the robot to perform a criminal act by remaining passive and not preventing the robot from doing it can be a thinkable scenario as well.[329]

In this case, the question of the criminal liability of the operator for the actions of his robot arises. Concerning autonomous driving, the operator may be obliged to keep his hands on the steering wheel of the car and watch over it for the entire course of its operation. If the operator does not attend to his duty and the car causes an accident, a programming mistake and, thus a breach of duty of care of the programmer, might be the cause of the accident. Yet whether the programmer can be accused of the criminal act in terms of foreseeability, avoidability or causality is questionable, since the operator also breached his duty of care. In order to determine who is criminally liable, the question needs to be answered whether this incorrect programming, in accordance with the ordinary course of events and experiences of life, was capable of causing the criminal act or whether the operator is to be held liable for breaching his duty of care?[330]

Further, the operator may also be held criminally liable for the actions of his robot, based on negligence, if he omits to prevent the robot from performing a criminal act.[331] According to German and Swiss criminal law, omission, with the specific legal exceptions,[332] is seen as a criminal act if the perpetrator had an obligation to act (the so-called guarantor's obligation).[333] A further prerequisite, in the case of omission, additional to the general prerequisites of negligence, is the guarantor status of the user.[334] Such an obligation to act can arise from legal

[329] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 178.
[330] Ibid.
[331] Ibid.
[332] In German criminal law, the so-called 'genuine criminal offences by omission', e.g. § 138, 323c StGB.
[333] Guarantor's obligation: Art. 11 paragraph 2 StGB (Swiss), § 13StGB (German).
[334] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique*

or contractual provisions, risk-bearing communities entered into voluntarily, or through the creation of risk.[335] In terms of robotics, a guarantor status might be established through the creation of a risk, as technologies tend to show increased risks at least in their early stages. There is a possibility that law makers will define duties for the operators of robots, which makes it thinkable that a guarantor's status may be established.[336]

9.5. Intentional Commission Using Robots as a Means of Crime

While the above-mentioned scenarios include cases where damages/injuries incurred unintentionally, or happened accidentally, there may also be cases where the programmer, producer or operator of the robot intentionally commits a criminal offence by using the robot's actions or help. In these cases, the question arises again as to how far the deeds of the robot can be attributed to the programmer, user or operator of the machine. A thinkable scenario might be the abuse of military systems, such as military robots and drones etc., or hacking bots in terms of cyber-crime.[337]

It is very likely that human beings use machines to knowingly harm others. For instance, the possibility of controlling a machine remotely allows humans to commit criminal offences without being on-site, without any risk of being directly caught or harmed while doing so. Perpetrators could, for example, commit a bank robbery with a remote-controlled robot.[338] In the future, committing such crimes may be much easier, particularly because, as a result of the possibilities offered by the internet, the perpetrator has the opportunity to commit the deed from any place in the world, making it much harder for him to be caught subsequently.[339]

Actuelle. 02/2017, p. 178.
[335] Art. 11 paragraph 2 subparagraphs a - d StGB (Swiss).
[336] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 179.
[337] Ibid.
[338] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 13.
[339] Neuhäuser, C. *Roboter und moralische Verantwortung.* In. Hilgendorf, E. *Robotik und Recht im Kontext von Recht und Moral,* 1st ed., vol. 3, Baden-Baden, Nomos, 2014, p. 270.

While criminal liability of remote-controlled robots is easy to clarify, criminal liability within the use of autonomous machines is much more complex to determine. Using an autonomous machine which is able to make decisions without the intervention or help of human beings bears significant risks towards the user or third parties. There are several reasons for this. On the one hand, AI systems nowadays have a comparatively high error rate.[340] On the other hand, they lack the qualities of compassion, creativity, and the ability to assess complex situations, leaving them unable to question the proportionality of their actions. However, these characteristics play an important role between human interactions, creativity and empathy.[341]

If a person commits a felony with the help of another person, criminal law differentiates between the case where the perpetrator knew about the wrongdoing and conformed with it, and the case of an unsuspecting perpetrator. In the first case, the second person could be an accomplice, a principle offender instigated by the first person or an assistant (according to German and Swiss Law).[342] If the person executing the wrong deed does not know that he is committing a criminal offence, the person might classify as an intermediary and the person being instigated might be an indirect perpetrator.[343] Given that robots, as far as their current development is concerned, are not able to understand the concept of criminal wrongdoing, they do not possess any kind of personhood and usually only follow certain predefined commands from their respective programmer or user, which means the first scenario must be excluded from consideration. Thus, robots cannot be accomplices, cannot be instigated and cannot be assistants in a felony, as long as robots are proven not to possess the capacity of will and, thus, criminal responsibility.[344] The last issue in particular is connected to another

[340] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 14.
[341] Becker, M. *Autonome Waffen – Wie Roboter den Krieg menschlicher machen sollen.* 2014, https://www.spiegel.de/wissenschaft/technik/autonome-waffen-roboter-sollen-krieg-ethischer-machen-a-977614.html.
[342] Donatsch, A./Tag, B. *Strafrecht I: Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, pp. 153 ff.
[343] Ibid., pp. 188 ff.
[344] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 179.

problem, debated in both law and ethics, namely: the legal grey area of liability.[345]

Even though robots are, at least today, not considered as persons in terms of criminal law, the concept of perpetration of an offence using an innocent agent might be considerable in an analogous application.[346] This is called indirect perpetration. An indirect perpetrator is somebody who uses another person as his will-less or at least not intentionally acting instrumentality to achieve the criminal act through the other persons actions.[347] Thus, the one performing the wrongful act does not act intentionally and does not even need to be capable of criminal responsibility, as is the case with children.[348] It would be possible to consider robots not only as instrumentality but also as will-less instrumentality analogue to the human used by the indirect perpetrator. Therefore, the robot would remain unpunished and its deeds would be attributed to the indirect perpetrator.[349] However, this alternative is only a doctrinaöl idea, as criminal law already applies the practice to attribute the deeds of a robot to the programmer or user, through causality within the course of the final doctrine of acting (in German and Swiss law). Still, there is a human being behind every deed of a robot, as it is a human being which causes the deeds of robots through their programming and giving it a specific command, which provides the machine with the impulse to act. The command is then adequately causal to the deed of the robot, which then leads to a criminal act. Not every case will make it that easy to prove the causality, or the human acting behind the robots' deeds to perform the appropriated deed. Nevertheless, the doctrine of criminal law seems to be adequate to cover such cases already.[350]

---

[345] Matthias, A. *Automaten als Träger von Rechten*. 2nd ed. Berlin. Logos Verlag Berlin, 2010, pp. 13 ff.
[346] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 179.
[347] Donatsch, A./Tag, B. *Strafrecht I: Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, pp. 188 ff.; BGE 77 IV 88 E. 1; BGE 138 IV 70 E. 1.4.
[348] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 179.
[349] Ibid.
[350] Ibid.; for more about the final doctrine of acting: see Eisele, J. In. Schönke, A./Schröder, H. *Strafgesetzbuch.* 30th ed. Munich, C.H.Beck, 2019, section 15 StGB, recital 28 ff.

9.6. Moral Dilemma Situations

In a subcategory of intentional act committed by the programmer (or user), there is also the question of to what extent a programmer can be held responsible for the decisions and subsequent actions of a robot in which the robot commits a criminal act in order to avoid another legally prohibited outcome (so-called moral dilemma).[351] In this context, the frequently asked question concerning the programmers choice whether an autonomous car will, in the case of an unavoidable collision, rather hit a wall which leads to the death of the passengers, or if it will decide to hit the pedestrian in front of the vehicle. In this case, the programmer has to make a prior choice of how to program the machine. Therefore, criminal responsibility due to negligence is no possibility here.[352]

In criminal law, situations in which criminal actions must be undertaken to protect higher legal interests from immediate danger are defined under the legal justification of necessity.[353] Thus, an act, such as destroying a person's property, may be justified and not punishable if it is committed to protect higher-order interests, such as the bodily health of human beings. In cases of equivalent interests, as demonstrated in the above-mentioned example, necessity as justification is not applicable.[354] In case of a moral dilemma, such a weighting does not seem to be an appropriate solution. The programmer might prevent himself of being criminally responsible through a justifying conflict between values and duties,[355] as it is about two equivalent legally protected rights. Here the perpetrator is not capable of programming the car in a way that, in the case of an accident, the car is able rescue the lives of both the passenger and the pedestrian. Further,

---

[351] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 179; for more information, see Wohlers, W. *Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen.* In. *Basler juristische Mitteilungen.* 3/2016, pp. 113 ff.

[352] Ibid., p. 180; see also Berger, M. B. *Autonomes Fahren - Gedanken und Lösungsvorschläge*. In. *Zeitschrift für Haftung und Versicherung.* 03/2016, pp. 289 – 295.

[353] Section 34 StGB (German); Art. 18 paragraph 1 StGB (Swiss).

[354] Donatsch, A./Tag, B. *Strafrecht I: Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, pp. 234 ff.

[355] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 180; for more information, see Donatsch, A./Tag, B. *Strafrecht I Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, pp. 2 ff.; Seelmann, K. In. Niggli, M. A./Wiprächtiger, H. *Basler Kommentar Strafrecht I.* 3rd edition, Basel, Helbing & Lichtenhahn, 2013, Vor. Art. 18 N StGB, recital 16 ff.

due to a mitigatory act in a situation of necessity,[356] or necessity as defence respectively,[357] the programmer might be safe from criminal responsibility if he can prove that due to a moral conflict, he decided to program the machine to decide that specific way within a dilemma situation.[358]

Autonomous driving further gives rise to ethical and legal issues regarding collision avoidance systems. There is no reason to only refer to the risks of autonomous driving. The manifold possibilities which this new technology provides, such as gains in mobility for elderly and disabled people, better road safety, environmental protection, less energy consumption, and the overall ease of transport deserve to be rated positively.[359]

Most likely collision avoidance systems will contribute towards reducing the number of accidents in road traffic. Nevertheless, these systems will cause accidents as well, including those which would have not happened in the absence of evasive manoeuvres, directed by the system. However, collisions occur in road traffic on a daily basis, some of them resulting in injuries or even death. Human drivers, facing complicated, accident situations are often overwhelmed with the situation and thereby not able to make correct and well thought decisions.[360]

Some legal systems strictly apply the principle of lesser evil in situations where the autonomous vehicle cannot avoid violating legal interests. Section 34 of the German Criminal Code rules governing necessity in the criminal law. According to this, protected interests must significantly outweigh the interests of the 'victims'. In practice this means that the damage must be kept as low as possible. In this sense, it is justified to damage a person's property to avoid killing another. The same rules apply in the so-called trolley dilemma. If an

[356] Art. 18 paragraph 2 StGB (Swiss).
[357] Section 35 StGB (German).
[358] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 180.
[359] Hilgendorf, E. *Automatisiertes Fahren und Recht.* In. 53. *Deutscher Verkehrsgerichtstag 2015*, Köln, Luchterland, 2015, pp. 57 ff.
[360] Hilgendorf, E. *The Dilemma of autonomous driving: Reflections on the moral and legal treatment of autonomic collision avoidance systems.* In. Hilgendorf, E./Feldle, J. *Digitization and the Law.* 1st ed., vol. 15, Baden-Baden, Nomos, 2018, p. 59.

autonomous vehicle gets into a situation where it would kill at least one person, the vehicle is allowed to take a manoeuvre which damages other people's property but saves the people which would otherwise be severely, or even fatally injured. According to the hierarchy of values of the German legal system, the lives of human beings always have a higher value than things.[361] However, the trolley dilemma mostly implies a decision of whose life is to be protected over another's. Following the principle of lesser evil is extremely difficult. This situation is one of the fundamental legal and ethical problems that any collision avoidance system will probably be confronted with. How should the vehicle react? Which person's life should be saved? Who should be killed?

The German Federal Constitutional Court made a highly relevant decision[362] concerning section 14 of the Aviation Security Act.[363] In this case a commercial airplane, transporting passengers was hijacked and brought under control by terrorists, intending to use the airplane as a weapon. The question is whether law should allow an airplane, full of innocent people to be shot down to save the lives of other innocent human beings. The idea of shooting down commercial airplanes for this reason was denied by the court. The court saw a breach against Art. 1 and 2 of the German Constitution.[364] Further, the German government is not permitted to protect its citizens by killing a minority – in this case the crew and passengers of the airplane – intentionally. Weighing life with life by the people who may possibly be killed (from shooting the airplane), and how many people might be killed (by letting the terrorists complete their intended action) is unlawful. The government is not permitted to kill human beings simply because the number of casualties would be less than the number of those saved.[365]

In 2016, the German minister of transport commissioned the ethics committee to develop a code of ethics for autonomous driving. The ethics

[361] Hilgendorf, E. *The Dilemma of autonomous driving: Reflections on the moral and legal treatment of autonomic collision avoidance systems.* In. Hilgendorf, E./Feldle, J. *Digitization and the Law.* 1st ed., vol. 15, Baden-Baden, Nomos, 2018, p. 60.
[362] BVerfG, 15.02.2006 – 1 BvR 357/05, recitals 1-156, NJW 2006, 751.
[363] Aviation Security Act of January 2005.
[364] BVerfG, 15.02.2006 – 1 BvR 357/05, headnotes 2, 3, NJW 2006, 751.
[365] Ibid., recital 38, NJW 2006, 751.

committee was successful in finding a solution which respects human rights.[366] The code dictates that, within accident-situations, any classification of potential victims, based on characteristics such as age, sex, physical or psychological condition, are strongly prohibited. Counterbalancing victims is prohibited as well, which means that it is not allowed to sacrifice persons so as to rescue others in emergency situations.[367]

Since antiquity, similar dilemma cases in which people were to be sacrificed to save other people's lives has been discussed in several scientific fields, such as law and, in particular, philosophy. Despite the sheer amount of literature that has culminated over the centuries, there is still no definite answer to such cases and the decision of the German Federal Court is unlikely to be the final answer in the ages long debate about balancing lives against other lives.[368] The principles in criminal law of prohibiting the quantification and qualification of human life and the incomparability of the value of life, are among the most frequently highlighted, yet rarely verified basic convictions of criminal law practice and criminal legal science.[369] Justifying the killing of people in order to save 'more' people's lives means that, in a variation of the trolley case, in which the autonomous vehicle has to decide to kill either one person or two should always opt for the action leading to a lower count of injuries and deaths. However, following the German Federal Court's decision, programming AI to act that way in emergency situations would not be permitted. However, in another decision, the same court faced the question of whether the person giving the command to shoot down an airplane to save other people's lives should be held criminally liable for the decision. The court decided to leave this question open unanswered.[370] The question of whether the programmer or producer of an autonomous vehicle could be subject to criminal liability for programming the

[366] Tafani, D. *Dilemmata der Maschinen. Künstliche Intelligenz, Ethik und Recht.* In. *Jahrbuch der Juristischen Zeitgeschichte,* 20(1), 2019, p. 311.
[367] Ethik-Kommission, *Automatisiertes und Vernetztes Fahren.* Bundesministerium für Verkehr und digitale Infrastruktur, 2017, p. 11.
[368] Hilgendorf, E. *The Dilemma of autonomous driving: Reflections on the moral and legal treatment of autonomic collision avoidance systems.* In. Hilgendorf, E./Feldle, J. *Digitization and the Law.* 1st ed., vol. 15, Baden-Baden, Nomos, 2018, p. 62.
[369] Schnieder, G. In. Joecks, W./Miebach, K. *Münchener Kommentar zum StGB*. 3rd, vol. 4, Munich, C.H.Beck, 2017, Vor. § 211, recital 29.
[370] BVerfG, 15.02.2006 – 1 BvR 357/05, recital 157, NJW 2006, 751.

AI to prioritize one life over another is of significant interest. It is conceivable that they might not be held accountable for such actions in certain circumstances. In situations where there is no chance of not killing at least one person, would it be ethically and legally more justifiable to not take the measure, which saves the most lives among all alternatives? A quantification seems unavoidable here. If we turn the situation around and take an autonomous car which is programmed to take the exact measure which kills the most people in emergency situations, most people will instinctively say that this is the most immoral alternative, and almost nobody would accept or even buy such a car. Most people will demand the number of victims to be kept as low as possible, thus the machine should save as many lives as it can. This means that we cannot ultimately discard the idea of quantifying the victims when the vehicle has to set off human lives against each other. There is a high probability that this outcome will align with the prevailing social ethics, which are commonly referred to as the moral intuitions of the majority.[371] In situations where at least one person will be killed, how else should the lowest number of people killed be achieved?

It is perhaps instructive to take the example of US-Airways-Flight 1549 in 2009, where the pilot had to conduct an emergency landing in New York. When the engines of the airplane stopped working due to bird strikes, the pilot was in a similar situation. He had to decide where to emergency land the plane as no airport was in range of the defect machine. This situation is well-suited as an example because the pilot had to make the decision which he thought was ethically the most correct: he navigated the plane over the Hudson River to make an emergency landing on water. The crucial factor was that he had to decide between trying to emergency land the plane somewhere in the city, perhaps on a wide road, a decision that would very likely lead to the death of the passengers, and everyone close to the landing plane. Landing the plane on the river, would most probably cause the death of the plane's passengers only. The pilot decided to opt for the solution which will cause the least number of deaths in a worst-case scenario. The pilot's decision is widely accepted as the correct one. Then,

[371] Hilgendorf, E. *The Dilemma of autonomous driving: Reflections on the moral and legal treatment of autonomic collision avoidance systems.* In. Hilgendorf, E./Feldle, J. *Digitization and the Law.* 1st ed., vol. 15, Baden-Baden, Nomos, 2018, p. 64.

if the decision with the lowest number of deaths, conducted by the pilot, is accepted as the right one, why should AI react differently, and probably cause criminal liability for acting exactly in accordance with this? However, this example delivers no answer for cases in which at least one person is killed, no matter how the machine or pilot decides. The other question is whether there are any differences in the expectations of human and AI. Human beings tend to become nervous in stressful situations. Especially, when an accident seems very likely to happen, panic reactions are highly common as well. It cannot be expected from human beings, with individual experiences, to make the ethically perfect decision, especially not under such pressure. AI, however, is not nervous and cannot panic. Further, it is unclear how public transport should react in such situations. If an autonomous driving bus (or tram which are in use already), filled with people must make such a decision, who should be sacrificed, looking at the perspective of sheer numbers?

There are more circumstances which may influence the expectations of society on the one hand, and criminal liability on the other. Does it make any ethical and legal difference if the autonomous vehicle is publicly or privately owned? As mentioned above, private owners will most likely expect their property to be 'loyal' towards their owners and, thus, to protect their lives more than those of other pedestrians and drivers. Publicly-owned autonomous vehicles may not have such obligations, which may change moral precepts. Some scholars argue that private owners of autonomous vehicles should carry more of the risk, as they are the ones introducing the potentially dangerous machine into public spaces.[372]

10. Criminal Responsibility of the Robots

As mentioned above, holding the robot as criminally responsible is not a potential option considering the recent technological development of AI. The idea of holding a robot criminally responsible for its deeds still fails due to the rather traditional understanding of guilt in criminal law. The traditional

[372] Lin, P. *Why Ethics Matters for Autonomous Cars.* In. Maurer, M. et al. *Autonomes Fahren.* Berlin Heidelberg, Springer Vieweg, 2015, p. 80.

understanding of guilt, or crime, is based whether the perpetrator can be blamed for the specific criminal act. The perpetrator must have had the possibility to act differently in the specific situation.[373] According to the German Federal Court of Justice, the reason to blame a perpetrator is the fact that human beings are able, based on free, responsible and moral autonomy, to make the decision of acting in a lawful manner and, thus, equally capable of deciding against acting lawfully.[374]

However, the idea of the German Federal Court of Justice basing criminal acts on human will and the decision between acting lawfully or unlawfully, has been relativised to a certain extent, by the concept of corporate criminal law, as applied by a variety of nations worldwide. While Germany is one of the last jurisdictions not to apply corporate criminal law, the USA, for example, recognises a kind of criminal liability, independent from guilt through so-called strict liability and vicarious liability.[375] Furthermore, Swiss federal law on administrative criminal law recognises the concept of criminal liability without guilt under certain circumstances. Here, criminal law already acknowledges a criminal liability without fault and imposes a fine for minor offenses. In case of disproportionate investigative effort, the prosecution of the natural person is waived, and the company is punished in their place. These concepts demonstrate how, in certain jurisdictions, the conventional interpretation of guilt has diminished in importance, with a more pragmatic alternative taking its place in these particular legal frameworks. The theoretical challenges/attacks of the traditional understanding of guilt are much more important than the pragmatic limitations of the traditional principle of guilt. It does not matter whether these challenges/attacks are influenced by the dispute concerning the freedom of will or from the dispute on the rational, and functional doctrine of criminal law. These disputes may gain

[373] Donatsch, A./Tag, B. *Strafrecht I: Verbrechenslehre.* Zürich, Schulthess Juristische Medien, 2013, p. 271.
[374] BGH, 18.03.1952 – GSSt 2/51, recital 12, – NJW 1952, 593.
[375] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 180.

importance again through the discussions about a possible responsibility of robots.[376]

It is clear that a functional definition of guilt or criminal responsibility would allow the existence of a guilty robot, if society develops in a way which allows robots to possess the personhood, going along with certain abilities and freedom of decision. The doctrine of criminal responsibility does not base the idea on naturalistic conceptions, but on the social purpose of the assignment of guilt in the course of a specific purpose of punishment.[377]

Moreover, the question concerning criminal responsibilities, or guilt of robots gives rise to the question of what purpose criminal law serves. Why should robots be punished? The purpose for humans is clear: punishment shall prevent the perpetrator from committing other criminal offences. This probably would not make any sense with robots as they currently exist, as they are pre-programmed and have no real freedom of decision-making. However, over the long term, the possibility that in the future, robots become increasingly autonomous and capable of learning, leading them to understand punishment as negative conditioning and to adjust their behaviour, cannot be excluded.[378] Already today, robots which feel pain does not seem to be impossible. German researchers are developing robots to feel pain and to react to these feelings.[379] Another function of criminal law, next to deterrence of perpetrators, on the one hand is the general deterrence of potential perpetrators and, on the other, the affirmation, and stabilisation of laws towards society. Thus, punishing robots can definitely make sense, as long as this robot is recognised as a person, bringing it into the scope of normative expectations. Independent from what effect the specific punishment on the perpetrator may have, its purpose may lie in the attribution of

[376] Markwalder, N./Simmler, M. *Roboterstrafrecht - Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 181; for more information, see Simmler, M./Markwalder, N. *Roboter in derVerantwortung? - Zur Neuauflage der Debatte um, den funktionalen Schuldbegriff.* In. *Zeitschrift für die gesamte Strafrechtswissenschaft.* 129(1), 2017, pp. 20 ff.
[377] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 181.
[378] Ibid.
[379] For more information, see Kuehn, J./Haddadin, S. *An Artificial Robot Nervous System To Teach Robots How To Feel Pain And Reflexively React To Potentially Damaging Contacts.* In. *IEEE Robotics and Automation Letters.* (2)1, 2017, pp. 72-79.

criminal responsibility, in order to maintain the expectations and norms of society. According to the actual development of AI, such an attribution of criminal responsibility is still unthinkable. Thus, according to modern state-of-the-art, holding machines criminally liable is not yet possible. However, it remains within the realms of possibility that there will be changes concerning humanoid robots, forcing criminal law and the criminal law doctrines to adapt to these developments.[380]

10.1. Digression to German Constitutional Law

According to Arts. 1 and 2 of the German constitution, free will and dignity of human beings is the highest legally protected right. Scholars describe the normative statement of the subjective constitutional law as follows: every human is human by virtue of his spirit, which contrasts him from an impersonal nature, and from his own ability to make decisions this enables him to become self-aware, to determine for himself and to influence his environment.[381] Consequently, it is crucial for the free will of human beings that they are capable of estimating the consequences of their acts, to make a rational decision based on rational thoughts. Another essential element of rational behaviour is the control over impulsive acts.[382] According to the German constitution, especially Art. 19 paragraph 2, the fundamental rights are the basis of its legal system. The German constitution places an emphasis on the rights of dignity and the freedom to act of human beings as indispensable and irrevocable fundamental rights. The relatedness of fundamental rights to natural persons shall not be rendered meaningless by other laws. In this sense, the German Federal Constitutional Court found that, according to Art. 103 paragraph 2 of the constitution, criminal liability always requires wrongful acts. This principle is rooted in Arts. 1 and 2.[383] According to Art. 19 paragraph 3 of the German Constitution, these rules apply to legal persons if these rules are "in so far as applicable by their nature." This limitation

---

[380] Markwalder, N./Simmler, M. *Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle.* 02/2017, p. 181.

[381] Antoni, M. In: Hömig, D./Wolff, H. A. *Grundgesetz für die Bundesrepublik Deutschland.* 13th ed. Baden-Baden, Nomos, 2022, Art. 1, recital 4.

[382] Cording, C./Roth, G. *Zivilrechtliche Verantwortlichkeit und Neurobiologie - ein Widerspruch?* In. *Neue Juristische Wochenschrift.* 68(1-2), 2015, pp. 26 ff.

[383] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 145.

means that fundamental rights, such as Arts. 1 and 2 are not applicable in this case.[384] Perpetrators and participants can only be natural persons. This implies that legal persons cannot be perpetrators according to criminal law, hence they cannot be sentenced to imprisonment.[385] Some scholars describe this as such: subject to German criminal laws is the human behaviour, meaning the activity, or passivity, which underlies the will of human beings.[386]

Historically, since more than a hundred years, the legal doctrine has said that law is a binding order of humans' coexistence, based on a collective's free will.[387] The rules concerning legal capacity (Section 104 ff Civil Code) clarify that a binding declaration of intent may only be existent, if a human being possesses cognitive judgemental capacities.[388]

For these reasons alone, German law provides no possibility in favour of a free will based on technology, and thus legal capacity of autonomous software agents.[389] Even if intelligent software agents possess problem solving capabilities similar to that of human beings, the German constitution would still not grant them legal personality with legal capacity.[390]

10.2. Strong AI and Criminal Liability

10.2.1. Problems of Strong AI in Criminal Law

10.2.1.1. Intention

Traditionally, the term mens rea has been used to imply the fault requirements, which are intention, recklessness towards a specific outcome, and knowledge of a certain circumstance. Intention plays the key role in serious crimes and sometimes constitutes the essence of an offence. In some cases, law

[384] Antoni, M. In: Hömig, D./Wolff, H. A. *Grundgesetz für die Bundesrepublik Deutschland.* 13th ed. Baden-Baden, Nomos, 2022, Art. 19, recital 5-10.
[385] Lackner, K./Kühl, K. In. Lackner, K./Kühl, K. *Strafgesetzbuch.* 28th ed. Munich, C.H. Beck, 2014, § 14, recital 1a.
[386] See Welzel, H. *Das deutsche Strafrecht*. 11th ed. Berlin, De Gruyter, 1969, pp. 33 ff.
[387] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 145.
[388] Brox, H./Walker, W-D. *Allgemeiner Teil des BGB*. 34th ed. Munich, Verlag Franz Vahlen, 2010, § 12, recital 259.
[389] Müller-Hengstenberg, D. D./Kirn, S. *Rechtliche Risiken autonomer und vernetzter Systeme - Eine Herausforderung.* Berlin, De Gruyter Oldenbourg, 2016, p. 145.
[390] Cornelius, K. *Vertragsabschluss durch autonome elektronische Agenten.* In. *Multimedia und Recht*, 5(6), 2002, pp. 355

grades offences with the help of intention such as manslaughter and murder.[391] For human beings it is seen as normal to do things with a certain intention in mind. However, in criminal law the question of what intention the perpetrator had when he or she committed the specific offence is irrelevant. The law rather focuses on the fact of the presence or absence of intention.[392] Thus, the question is not whether what specific intention the perpetrator had but rather if he generally had any intention when he or she committed the offence.

The particular intent is a requirement, and arises from the definition of the offence. It is essential to keep in mind, that there is the 'easy' possibility to say that the perpetrator pulled the gun's trigger intentionally but did not mean to kill the person at whom he directed the gun. Thus, in a case of murder the focus lies on the presence or absence of an intention to kill the victim. The fact that the perpetrator acted accidentally or intentionally might be an important fact in the case but, ultimately, the intention required by criminal law, is that the offender intended to kill the victim.[393] However, the core of intention is aim, objective or purpose. Regardless of any alternative meanings of aim, a person definitely acts intentionally if killing the person is either the goal of the action, its objective, or purpose, with the victim's death as result.[394] However, there are certain concerns about the term 'purpose' in terms of confusion among the purpose of the person's act with an ulterior motive on the one hand and, on the other, with the purpose a person's act to bring about an immediate result.[395] In the Mohan Case[396] the Canadian Court of Appeal held that the crime of attempt, oblique intention or recklessness, did not suffice, as direct intention needs to be proven:[397] according to James LJ: "[...] it is well established law that intent (mens rea) is an essential ingredient of the offence of attempt […] What is the meaning of 'intention' when that word is used to describe the mens rea in attempt? […] the word means what is often referred to as 'specific intent' and can be defined as "a decision to bring

[391] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 168.
[392] Ibid., p. 169.
[393] Ibid.
[394] Ibid.
[395] The Law Commission. *A new Homicide Act for England and Wales?* Rectials 4.36-37.
[396] R V. Mohan [1975] 2 All ER 193 p. 198.
[397] Clarkson, C. M. V./ Hill, J. *The Conflict of Laws.* 3rd ed. Oxford, New York, Oxford University Press 2006. 1976, p. 63.

about a certain consequence or as the 'aim'."[398] As pointed out, the Mohan case involved an attempted crime and shows that intention is thought to be essential to attempt, as "one cannot be said to attempt to produce a result unless one intends to produce it." To some extent, the decision of the Mohan case states that the core of the intention concept means acting in order to bring about a result.[399]

It must be kept in mind that, at least, Courts which apply the English common law, do not adhere to one single definition of the term intention.[400] As a reaction this is commonly treated as evidence of the court's behaviour. Courts seem to decide towards the desired result, and afterwards define the law in the way it may achieve it. However, this statement should be treated very carefully, as the evidence is limited to a few decisions of appeal courts, and, thus probably does not reflect the actual practice of the criminal courts.[401] The reason why judges have adopted a variety of different meanings of the term intention is mainly that the standard definition, combined with the set of possible defences to liability, sometimes may be insufficient in capturing moral distinctions which were thought to be of importance.[402] When it comes to the point that a criminal court has to decide over the guilt of AI, it makes sense that the deciding court might retain its 'freedom' in decision-making, in terms of interpreting laws. This way, courts might have the possibility to make the respective law suit cases in which mostly, or entirely AI is involved. If a robot points a loaded gun to the head of a human being und pulls the trigger, it is clear that the person will most likely be killed. A robot with strong AI is most probably trained and informed about such devices (weapons) and their properties. It is not hard to imagine that such a robot knows very well of the outcome of shooting a person in the head with a gun – why else should a robot point a gun towards somebody? Applying the above-mentioned arguments would probably open up enough space for judges to use their freedom in applying laws in accordance to the respective case,

[398] R V. Mohan [1975] 2 All ER 193 p. 198.
[399] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 170.
[400] For more information see Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 172-174.
[401] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 174.
[402] Ibid., p. 175.

making any deeper proof of mens rea superfluous, since the robot most probably knew what it had done. At this point even the robot's intention is no longer relevant within the decision-making.

10.2.1.2. The Principle of Mens Rea

Mens rea plays a major role in criminal liability and is expressed by the requirement of the mental element in a criminal offence. It is a general requirement for all types of criminal liability.

The application of mens rea in the (English common law) courts show that mens rea need not, necessarily, have to do with an evil mind, moral fault and knowledge concerning the wrongfulness of the act.[403] It is not considered as a defence if the perpetrator was not morally at fault.[404] It is further no defence if the defendant was ignorant towards the fact that his conduct was an offence, if he did not consider that his conduct was immoral or did not know whether his conduct is seen as immoral by society. Additionally, it is irrelevant for criminal law if the perpetrator had a good or bad motive.[405] Generally, mens rea is thought to mean 'guilty mind', however it actually refers to any state of mind a person must be proved to have had in accordance with the definition of the crime charged. It further comprises other fault elements which are independent of a specific state of mind, for example, negligence and objective recklessness. This is important to do so, as using the term mens rea in its narrow sense would mean the exclusion of important indicators of blame, for example, negligence. Due to this, the UK Law Commission uses the term 'fault element', which might be more appropriate.[406]

As mentioned above, the strict sense of mens rea implies any state of mind one is required to have in accordance with the definition of the crime charged. Usually, this goes hand in hand with proving that the perpetrator was at least aware that he might perform the actus reus of the crime, if he continues with the planned conduct. On occasion, it may entail more than that, such as

[403] Card, R. *Card Cross and Jones Criminal Law.* 17th ed. Oxford, Oxford University Press, 2006, p. 88.
[404] Yip Chiu Chueng v. R. [1995] 1 AC 111.
[405] Card, R. *Card Cross and Jones Criminal Law.* 17th ed. Oxford, Oxford University Press, 2006, p. 88.
[406] Heaton, R. *Criminal Law Textbook.* 2nd ed. London. Oxford University Press, 2006, p. 47.

demonstrating his intention to commit the actus reus of the crime.[407] Similar to the decision of the German Federal Court, the perpetrator may recognise the fact that he might perform the actus reus of the crime and nevertheless continues with the act. He chose to commit the crime, or chose to take the risk to commit it via his action. Thus, serious crimes such as violence against a person may require awareness of the possibility to harm somebody as a result of the conduct. However, mens rea in its strictest sense is not always provable due to a lack of awareness. Still, the lack of awareness is blameworthy when the perpetrator ought to have been aware of the possible outcome, thus committing the actus reus.[408] In this case, he is blamed for failing to think and act according to the ordinary standards of conduct.[409] Here, German criminal law applies a similar doctrine as well. Given that it is not always possible to foresee the details concerning the course of events precisely, deviations between the imagined and actual course of events do not exclude the existence of intent, if they are in the sphere of what can be known according the general experience of life, and where there is no other justifiable way to evaluate the act in question.[410] Thus, also German criminal law applies the foreseeability rule when mens rea is not exactly provable, for example in cases where the perpetrator was in error of causality of his act. The abovementioned approach[411] that has been applied, for example, in English criminal courts for a long time is considered in German criminal law as well. Older literature applied the 'Wahrscheinlichkeitstheorie', the 'theory of probability', where the perpetrator acted with conditioned intent when he assessed the outcome of his act not only as possible, but highly probable.[412] Although German legal scholars have largely abandoned this theory, it is experiencing a renaissance[413] in Germany and may be reconsidered in the future, particularly in cases involving AI technology.

Another doctrine which is mostly applied in German criminal courts is the 'Billigungstheorie'. It holds that the perpetrator acts with intent, if he accepts the

[407] Heaton, R. *Criminal Law Textbook.* 2nd ed. London. Oxford University Press, 2006, p. 47.
[408] Ibid., p.48.
[409] Ibid., p. 49.
[410] Rengier, R. *Strafrecht Allgemeiner Teil* 11th ed. Munich, C.H.Beck, 2019, § 15, recital 11.
[411] Referred to as 'Bedingter Vorsatz', which can be translated as 'conditional intent'.
[412] See Mayer, H. *Strafrecht, allgemeiner Teil.* Stuttgart, Kohlhammer, 1967, p. 121 ff.
[413] Frister, H. *Strafrecht Allgemeiner Teil* 9th ed. Munich: C.H.Beck, 2020, chapter 11, recital 26.

occurrence of the actus reus (or outcome) as possible and not completely improbable. This way, the perpetrator accepts the occurrence of the actus reus, or accepts the occurrence of the actus reus within his actions in order to fulfil his aspired goal, even if the actus reus may be undesired by him.[414]

An offence must contain a subjective mens rea requirement to alert the perpetrator that he is on the verge of violating the law, in order to satisfy rule of law standards. It is possible to interpret the principle of autonomy in a further sense arguing that the incidence and the degree of liability shall reflect the choices which the individual, or perpetrator, made. Mens rea holds that defendants shall only be held criminally liable in cases where the perpetrator intended or knowingly risked the events or consequences, thus, the perpetrator's conduct shall be deemed as criminal if they acted with awareness concerning their acts potential outcome.[415] Intention means 'aim' or 'purpose', hence a perpetrator intends a result if he acts in order to bring it about:[416] "it is the intention to carry out the crime that constitutes the necessary mens rea for the offense."[417]

German criminal law has a different approach to the meaning of intention, as it has three types of intent: intent in its strict meaning, called 'intent first degree'. Dolus directus, or "Intent Second Degree" and dolus eventualis, referred to as "Intent Third Degree." Intent First Degree is defined as the will tied to the deeds purpose. Here solely the conducts result is of relevance and the final goal is moot. Intent Second Degree is given in cases of intent where knowledge rules, thus, the defendant is aware of the future occurrence of a certain incidental consequence. Whether the defendant desires the consequence is insignificant. The requirement of the Intent Third Degree is that the defendant anticipates the outcome as a possibility, while accepting that the possible

[414] BGH, 04.11.1988, 1 StR 262/88, BGHSt 36, 1 p. 9 ff.; BGH, 22.03.2012, 4 StR 558/11, BGHSt 57, 183, pp. 183, 186.
[415] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 74.
[416] The Law Commission. *A new Homicide Act for England and Wales?* Rectial 4.6.
[417] Yip Chiu Chueng v. R. [1995] 1 AC 111.

outcome of his or her conduct might bring about the certain result, although the result is undesired.[418]

The essence of mens rea is to only impose criminal liability on persons sufficiently aware of their deeds, while being aware of the resulting consequences. It is based on the principle of autonomy which states that individuals are considered to be autonomous persons with the ability to choose among several alternatives of behaviour, considering that they are held liable based on their choices.[419] The German Court of Justice is of a similar opinion. In its decision it points out that guilt requires accusability. Accusability means that the defendant did not act lawfully, even though he could have decided to do so. The accusation is based on the person's capacity of acting according to his free will, responsibility and moral autonomy, which enables him to decide to act in lawful, moral manner, and to decide against taking unlawful action. This capacity requires awareness of right and wrong.[420]

The structure of the mental element requirement applies the fundamental principle of culpability in criminal law, nullum crimen sine lege. The principle of culpability rests on two main aspects: the positive and the negative. First, the positive aspect, or what should be in the offender's mind in order to impose criminal liability, relates to the mental element. The negative aspect describes what should not be in the offender's mind to impose criminal liability, and relates to the general defences.[421]

As an example: the mental element requires recklessness to impose criminal intent for physically harming a person. On the other hand, the negative aspect requires that the offender was not insane while harming the person. Part of the positive aspect of culpability is recklessness. The general defence of insanity represents the part of the negative aspect. In criminal law, the positive aspect of

[418] The Law Commission. *A new Homicide Act for England and Wales?* Rectial 4.10. For more information on intention see Sternberg-Lieben, D./Schuster, F. In Schönke, A./Schröder, H. *Strafgesetzbuch.* 29th ed. Munich, C.H. Beck, 2014, § 15, recitals 60 ff.
[419] Ashworth, A./Horder, J. *Principles of Criminal Law.* 7th ed. Oxford, Oxford University Press, 2013, p. 155.
[420] BGH, 18.03.1952 – GSSt 2/51, recital. 12 – NJW 1952, 593.
[421] Ashworth, A. *Principles of Criminal Law.* 5th ed. Oxford, Oxford University Press, 2006, 15 ff.

culpability is connected to mental processes within the commission of the criminal offence, showing two important aspects of cognition and volition.[422]

Cognition is defined as the individual's awareness of the factual reality.[423] There is no way a person may be aware of the future, and prophecy skills are not a subject of criminal law. Thus, criminal law cannot require a person to possess prophecy skills. Criminal law refers to cognition in a sense that either the offender is aware of a specific fact or not. Generally, criminal law does not accept partial awareness, and classifies it as unawareness.[424]

Volition is connected to the will of the individual, which is not subject to factual reality. Individuals may wish for the occurrence of unrealistic events. Due to the different levels of volition, it cannot be defined as binary. Volition has three different basic levels: first the positive level, which means the will of the occurrence of a specific event. The neutral, where one person is indifferent towards the occurrence of an event, and the negative, in which one person does not want a specific event to occur. There are intermediate levels between the basic levels of volition, for example, the rashness level. Here, the perpetrator does not want a specific event to occur, but takes unreasonable risk towards it. If that person absolutely did not want the specific event to happen, he would not have taken the unreasonable risk towards it.[425] A commonplace example for this is where a driver of a car refuses to stop at a red traffic light and hits a pedestrian while crossing the street, causing his death. The driver of the car did not intend or wish for the death/injury of the pedestrian. However, he took the unreasonable risk, which may fulfil an intermediate degree of volition. In the opposite conclusion, where the driver of the vehicle absolutely did not want to hit and kill the pedestrian, he would not have taken the unreasonable risk and performed the highly dangerous procedure of overtaking through a continuous line. The mental element requirement, deduced from the positive aspect of guilt in criminal law,

---

[422] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 68.
[423] Sullivan, G. R. In. Shute, S./Simester, A. *Criminal Law Theory: Doctrines of the General Part.* Oxford Monographs on Criminal Law and Justice, 2002, p. 214.
[424] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 68.
[425] Ibid.

is established by the combination of cognitive aspects. The majority of modern legislations recognise three primary types of the cognitive element. They are distinguished on the basis of cognitive factors, and are to be found in three different types: general intent, negligence and strict liability.[426]

The highest level of the mental element is general intent, known as mens rea, which requires complete cognition. It will play a major role regarding the question of punishing AI systems. Mens rea requires the perpetrator's full awareness of the factual reality and includes the examination of the subjective mind of the perpetrator. Negligence is defined as cognitive omission, where there is no requirement of awareness of the factual element on the side of the perpetrator. However, the perpetrator, based on the key facts of the specific case, could have been aware of it and should have been aware of it as well. The type of liability with the lowest threshold is strict liability. It presumes negligence on the basis of the factual situation, where the offender has the chance for rebuttal.[427]

The factual reality relevant in criminal law is that which is reflected by the factual element integral part. In criminal law, the factual reality that matters is determined by the factual element components. The perpetrator's perspective is limited to the conduct and circumstance components, which are relevant in the present. The results components, on the other hand, come into play in the future. As cognition is limited to the past and present, it can only pertain to conduct and circumstances.[428]

Against the fact that results occur in the future, the chance that the result may take place following from the germane performance is already existent in presence. Thus, cognition concerns conduct and circumstances, and further the results coming to pass. In the homicide case-example, the perpetrator aims a firearm at another person and pulls the trigger. The perpetrator is well aware of

[426] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 69.
[427] Ibid.
[428] Ibid.

his performance and also about its possible, or rather probable outcome: the victim's death.[429]

Volition is an insignificant element for negligence and strict liability. It can solely be associated with the mental element requirement of general intent. General intent, as mentioned above, comprises the full set of levels of will. The majority of legislations see general intent as the prescribed condition for the mental element. Therefore, criminal deeds of negligence and strict liability are required to clearly define the requirements pertinent in the certain case.[430]

If there is no such clear requirement mentioned, the criminal act qualifies as a general intent offence. The general intent offence is the default requirement. The relevant requirement, on one hand, can be fulfilled by the same form of mental element; it may, on the other hand, be met by a higher-level form as well, so that the mental element requirement of the offence is the minimal level of mental element required for the imposition of criminal liability.[431]Any lower level is not considered sufficient in order to impose criminal liability for the offence.[432]

Some legal systems distinguish between specific intent relating to purposes and motives, from intent having to do with the occurrence of a specific result. However, using the terms which are used by the different legal systems, many of them have in common that the required mental element component is the one reflecting the highest level of volition or positive will of the offender. It is further the one which does not relate to a given component of the factual element requirement, but to the purpose of the completion of the offence.[433]

Relating to the current structure of the mental element requirement, the minimal requirements to impose criminal liability is embodied by each specific offence. The fulfilment of the requirements is enough to impose criminal liability, making any additional psychological meanings unnecessary.

[429] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 69.
[430] Ibid.
[431] Ibid., p. 69 ff.
[432] Ibid., p. 70.
[433] Ibid., p. 72.

Consequently, an individual fulfilling the minimal requirements of the relevant offence is an offender, resulting in a possible imposition of criminal liability.[434]

10.2.2. The Existence of General Intent in AI

10.2.2.1. Cognitive Capacities of AI Systems

General intent contains a cognitive aspect. This however is based on awareness. Here the question arises whether AI has the capacity of awareness.[435] Given that the meaning of "awareness" can differ across scientific disciplines, such as psychology, theology, and law, among others, it is important to examine its legal definition to answer this question. Criminal law defines awareness as perception by senses of factual data and its understanding.[436] Data is processed in the brain as an internal process. Factual data, which the brain receives from the sensor organs, such as eyes, ears etc., is processed to a relevant general image of the factual data. This is defined as perception. However, most of the time the brain receives the input of many stimulations at the same time. The human brain has to focus in these cases on some of the stimulations and ignore the others to create an organised image of the factual data. This is called attention.[437] Attention plays a major role in autonomous driving for example. Human beings cannot focus on everything at once. AI, as it works very differently, does not recognise the idea of focus and, as long as it has the computing capacities, it is able to accord its full 'attention' to every piece of data it receives from its sensors. However, the brain accumulates all stimulations and creates the relevant general image. This process is done unconsciously, thus, human beings are not aware of it. The result, or the relevant general image is the conscious result. The human mind is aware of the result when the relevant general image is accepted. This process is what human awareness is based on. Being aware is the highest state of perception. Perception of the factual data by sense and its understanding results in the creation of the relevant general image. Creating the relevant general

[434] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 70.
[435] Ibid., p. 86.
[436] Lycan, W. G. *Mind and Cognition: A Reader.* Hoboken, New Jersey, Blackwell Publishers, 1990, pp. 3–13, cited: Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 86.
[437] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 86.

image is what is defined as awareness of the factual data. In a very technical perspective of the human body, the eyes function as nothing more than light sensors, delivering factual data to the brain. If the brain, in the next step, creates the relevant general image, the human is aware of what it is seeing.[438] To clarify, the human brain, not the eyes, serves as the organ responsible for sight in humans. The eyes act solely as sensors, transmitting data to the brain. It is only when the brain processes this information and creates a cohesive image that a person can be said to be aware of what they are seeing. In sum, there are two conditions which need to be fulfilled accumulatively for human beings to be considered aware of certain factual data. First, factual data need to be received by the senses. Second, the brain needs to process a relevant general image from the data it has received. In any given situation where at least one of these two requirements is missing, then the person is not considered to be aware. As the term knowledge is sometimes used to describe the cognitive aspect of general intent, the question arises whether there is any difference between knowledge and awareness. The answer is that criminal law does not differentiate between knowledge and awareness as they refer to the same idea of cognition.[439] Some judgments even define knowledge as awareness: "when knowledge of the existence of a particular fact is an element of an offense, such knowledge is established if a person is aware of a high probability of its existence, […]"[440] and "a violation of an OSHA regulation, either by act or omission, is 'wilful' if it is done knowingly and voluntarily, either in reckless disobedience of the regulation or in reckless disregard of the requirements of the regulation."[441]

As science could not undoubtedly prove the existence, or the scientific (neurologic etc.) basis of awareness, it is very hard to prove the offender's awareness in court without any doubt as it is required by criminal law. As a solution to this problem, criminal law has developed substitutes to circumvent the problems and solve this in the form of presumptions. These presumptions are applied in some situations and presume the existence of awareness. These

[438] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, pp. 87-88.
[439] Ibid., p. 88.
[440] United States v. Jewell, 532 F.2d 697 (9th Cir. 1976).
[441] United States v. Ladish Malting Co.,135 F.3d 484 (7th Cir.1998).

presumptions (of which two are major) are recognised in most legal systems. The first presumption is that of willful blindness, which functions as an alternative to being aware of one's conduct and the surrounding circumstances.[442] Wilful blindness is presumed if the offender suspected the fact and realised its probability, but still refrained to obtain the final confirmation as, in the event, he wanted to be able to deny knowledge.[443] The second is to presume awareness as a substitution for awareness of the possibility of the occurrence of the result. Now the question arises whether AI is capable of being aware of conduct, circumstances and the possibility of the result's occurrence. First, AI is capable of receiving the specific factual data by its sensors. These sensors are very comparable to human organs having such functions: eyes provide visual data as 'light sensors', and human ears provide auditory senses comparable with microphones etc. Our sensory organs provide us the factual data and transfer the data to our brains, which processes these data through the nerves (which are comparable with data transmission cables). In essence, there are hardly any differences to AI as it can have the same abilities through its sensory elements. This makes AI capable of receiving factual data as their sensors absorb light, and their microphones absorb sound etc., and transfer these data to their processor units. In fact, most modern sensors are much more accurate and sensitive to their respective capacity than their biological, human equivalent. Additionally, these sensors are sensitive to a broader extent of sound and light wavelengths, which means that these sensors are able to see colours and hear sounds of which human beings are not capable.[444] In consequence, AI can fulfil the 'first stage' of awareness and is even able to do it better than human beings. The next stage is to create the relevant general image from these data. As AI does not have a biological, but an artificial brain, one might ask the question whether 'real' intelligence, or awareness necessarily needs a biological brain to exist. While this question goes beyond the scope of this work, it raises an important issue: can processors or artificial 'brains' generate a relevant and

[442] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 89.
[443] Williams, G. *Criminal Law: The General Part.* 2nd ed. London, Stevens & Sons Ltd, 1961, § 57, p. 159.
[444] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 89.

comprehensive image from factual data? Human brains create the relevant general image through analysis of the factual data to use, transfer and integrate with other information and act in accordance with it, hence, understanding it.[445]

A very good example here are AI-based security robots. Their purpose is to defend (e.g. property) by identifying intruders and to call human police officers or armed forces etc. They also may stop intruders by themselves. Here, the cameras and sensors recognise the intruder by receiving the factual data and send them to the processor. The processors analyse the factual data and eventually identify the person either as a known person or an intruder. When analysing the factual data, the processor is not permitted to confuse the intruder with policemen or soldiers etc. and thus the processor identifies visual and audio changes. A further step is the assessment of the probabilities. If the probabilities do not allow for accurate identification, it conducts a vocal identification by making the person identify himself, for example, through a password. When the person in question answers, the microphones of the AI system receive the vocal sound of the person and the processor can now compare the vocal sound with the voices it has in its memory. Based on that factual data the machine can now make a decision, as the processor has generated the relevant general image from the factual data which the processor received through the sensors. This specific relevant general image enabled the machine to use, transfer and integrate the information with other information and made it act in accordance to it. At this point it is adequate to compare this with the behaviour of a human guard who, it is argued, would very probably act the same way. The guard first hears or sees the person in question. Based on the visual information, or the factual data the person receives from the senses, the brain processes the information and creates the relevant general image. Just like the robot, the human guard would compare the sound or voice and the visual data with the memories the guard has to identify the person in question. If the guard is not able to identify the person accurately, the guard would call for identification as well.[446]

[445] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 89.
[446] Ibid., p. 90.

Both the human guard and the robot guard made the relevant decision. Both understood the situation. At this point, some might say that the human being was aware of the relevant factual data, but why could this not be said about the robot guard as well? In conclusion, there is no reason to deny the awareness of the relevant factual data on the side of the robot. The internal process of the human guard and the robot were actually the same.[447] The only difference is the medium through which the information is processed: the human brain is biological, while the processor of the machine is made of silicon. On an even deeper level of comparison it must be mentioned that the carrier of the signals and, thus, the information is identical for both, biological brains and silicon processors: electronic signals. Ultimately, the functioning of the two does not differ as one might think.[448]

Another example is a machine deciding to kill a person with a gun. The AI of the machine will know what a gun is, how it looks, how it works and that it is designed to kill. When the machine picks up a gun, walks to the targeted person and points it to the person's head, the machine will very well know what will happen if it pulls the trigger. It knows very well that there is no law allowing this. There are plenty of other obvious signs, such as body language and things the person might say. The machine collects all the factual data, processes them and knows very well that he is already harming a person, just by pointing a gun towards him.

Consequently, in theory, AI systems may be capable to fulfil the second stage of awareness. The two stages examined above are the only relevant stages of the awareness process relevant for criminal law. This allows the final conclusion that machines are capable of fulfilling the requirements of awareness. At this point some might say that a number of the above-mentioned criteria of cognition to convincingly prove the AI machines capability of awareness are lacking. This might be correct in the broader sense of awareness, which philosophy, psychology, neurology and cognitive sciences require or attempt to prove.[449]

---

[447] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 91.
[448]Bridgeman, B. *Brains + programs = minds.* In. *Behavioral and Brain Sciences.* 3(3), 1980, p. 427.
[449] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer

Others might say that the above-mentioned criteria are not the deciding criteria to make an entity aware.

However, as concluded before, none of the sciences have been able to prove the existence of awareness, thus, criminal law has to find a way to circumvent the 'inevitability' of awareness and create possible ways to penalise perpetrators for their criminal deeds, without the actual proof of awareness.

The focus of criminal law should be on evaluating the culpability of AI technology, rather than delving into the broad interpretations of cognition found in fields such as psychology, philosophy, and cognitive sciences. Due to this, the only standards of awareness relevant to examine are those of criminal law, which makes all other standards, more or less irrelevant for the assessment in terms of criminal liability. This counts for both AI and human intelligence. Criminal law defines awareness due to the above-mentioned complications in a much narrower fashion than the other relevant sciences.[450]

Awareness is very hard to prove in the courts. This is equally so in criminal cases, where in particular it must be proven beyond any reasonable doubt. Therefore, criminal law has developed substitutes, as mentioned above, to circumvent this issue: the presumption of wilful blindness and awareness.[451]

10.2.2.2. Volition and AI

Volition is defined as the capability of humans and animals to conduct actions on the basis of their internal decisions and motivation, rather than their external stimulation.[452] In criminal law, volition or the volitive aspect of general intent is composed by three levels of will: intent, indifference and rashness.[453] Most modern legal systems refer to indifference and rashness as recklessness. Thus, the question arises whether AI has the capacity to consolidate such levels of will.[454] As these terms may have different meanings in other sciences, such as

---

Cham, 2015, p. 91.
[450] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 91.
[451] Ibid.
[452] Haggard, P./Lau, H. *What is volition? Experimental Brain Research.* 229(3), 2013, p. 285.
[453] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 93.
[454] Ibid., p. 94.

psychology, philosophy etc., the relevant terms here are the definitions in criminal law, as it is the question of criminal liability that shall be examined in this work.

According to criminal law, intent is the highest level of will. The aim of the conduct is the achievement of a specific purpose. Intent further is defined as the positive will, or as the will that a specific actual event occurs. Of course, higher levels of intent do exist. However, according to criminal law, intent is the highest imposition of will, possibly required to impose criminal liability, depending on the particular offence. Proving both intent and awareness are very difficult in cases concerning criminal law, even though both are internal processes of the human mind. The difference between them is that awareness relates to current facts whereas intent relates to a future factual situation. Awareness is based on rationality and is realistic, while intent is not necessarily rational or realistic. Criminal law often applies the foreseeability rule or dolus indirectus. This is the legal presumption with the specific purpose to prove that intent is existent. It presumes that the perpetrator wants a specific result to occur, if the offender has foreseen the result's occurrence in a sense of an event which is very likely to occur. As an example: a person points a loaded gun to another person's head, knowing that his death by shooting his head is a factual event which is very likely to occur. Thus, he pulls the trigger. If the offender argues that he did not wish for the death of the victim, the required intent is not fulfilled, and the court should acquit the offender. However, if the court applies the foreseeability rule presumption, it is presumed that the offender intended the results to occur. The shooter knew that the victim's death will be the outcome, or at least assessed the result as a very highly probable. Therefore, he is presumed to want the occurrence of the result.[455]

Now the question arises whether AI is capable to have intent in a relevant sense to criminal law. As said before, this is difficult to prove as it is an imprecise term, but the question whether AI may be capable of having intent should be answered the same way as criminal law answers with human beings: through the

[455] Shute, S. In. Shute, S./Simester, A. *Criminal Law Theory: Doctrines of the General Part.* Oxford Monographs on Criminal Law and Justice, 2002, pp. 182–187.

foreseeability presumption. Two conditions need to be fulfilled within this presumption. First, the occurrence of the result needs to be foreseen with a very high probability, and that the conduct has been done with awareness. Strong (and probably weak AI as well) are already able to assess the probability of occurring events. Computers made for playing chess analyse the game status. Based on the positions of the chess pieces the computer runs a variety of possible options to plan the next move. Further, they plan the reactions to possible reactions from the other player etc.[456] It is not only chess playing computers that do this. Computers running a strategy game make similar moves. They apply strategies to win the game and so do computers in other genres, such as ego shooters etc. The computer evaluates the likelihood of each option before selecting its subsequent action. Thus, if it was a human being, it would form the impression that the person has the intent to win the game. There would be no doubt. Nevertheless, the true intent of the person is not provable, even though his conducts fits the foreseeability presumption. AI behaves in a goal driven manner to win the game, just as the human player does. Therefore, it is said that the human player intends to win the game. The same might be said about playing AI systems as well. Analysing the courses of their conduct show that both match the foreseeability rule presumption.[457] Examining the options of conduct and making a decision in awareness to commit them, based on the assessment that the specific event will most probably result from the conduct, is considered as foreseeing the occurrence of the factual event. If the results and the conduct together form a criminal offence, it falls within the scope of criminal law. It makes no difference if it is a chess game with the purpose to win the game and to commit any other conduct with the purpose of the occurrence of a result. In the case where results, paired with the conduct constitute a criminal offence, it falls under the scope of criminal law.[458] In consequence, if an AI system undertakes an assessment of probabilities that a specific factual event, for example, the death of a person or winning a game etc., will be the result of its conduct with a very high probability, and chooses to commit the relevant conduct

[456] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 97.
[457] Ibid.
[458] Ibid.

accordingly, the AI system fulfils the conditions of the foreseeability rule presumption. Thus, the AI system is presumed to have the intention to make the results occur. In most cases, courts examine the offender's intent in this exact way, when the offender refuses to confess. The difference is that AI systems have much more precise abilities to assess probabilities than human beings. AI technology can have the ability to consolidate awareness to factual data. As the commission of conduct is considered as factual data, AI is able to fulfil both conditions for the foreseeability rule presumption, which is classified as an absolute legal presumption or preasumptio juris et de jure. Under these conditions, the conclusion of having intent is incontrovertible. However, one must keep in mind that the foreseeability rule presumption can only be based on strong AI. The more difficult requirement is the ability of AI systems to use the assessment of probabilities as a tool for their decision-making. Strong AI possesses such capabilities. Which results in the fact that strong AI has the capability to intend in a sense which falls under the scope of criminal law. Thus, AI is capable of fulfilling the intent requirement required by criminal law.[459] The definitions for intent and foreseeability in criminal law are much narrower than the definitions brought by other sciences such as psychology, philosophy and neurology etc. However, this is true for imposing criminal liability upon AI systems and human beings. Actual evidence for the intent of AI is based on the possibility of monitoring and, in particular, recording all activities of the software. Every step or component of the consolidation of the machine's intent and foreseeability is monitored and further recorded and saved by the AI software running the machine, as it is part of its activity. It further is a part of the machine's activity to assess probabilities and make the relevant decision in accordance with the given situation. Despite this, extracting and investigating the records of the machine might be superfluous. For criminal law, the proof through the foreseeability rule presumption provides direct evidence of the AI's criminal intent. Requirements in law in general represent only a minimum of condition to impose criminal liability. Due to this fact, strong AI or at least AI systems possessing the above-mentioned capabilities concerning foreseeability, can fulfil the requirements of the mental element of intent on the one hand and

[459] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 98.

recklessness on the other. However, AI is capable of directly fulfilling the requirement of recklessness as well.[460]

Decision-making is based on highly complicated processes. For AI technology, these processes are characterised by a large number of different factors that the machine needs to consider. It is similar with humans as well, but humans mostly ignore a part of factors and, thus, do not take them into consideration. Computers and AI technology can do the same. Computers, in particular, are programmed to exclude some factors. However, strong AI is capable of learning to ignore certain factors. Without excluding certain factors, decision-making might be impossible. This is based on machine learning, or learning from examples. The effectiveness of this type of learning rises with the number of examples analysed and can be compared to 'experience'.[461] However, excluding some factors might not be the correct way. If the hardware is able to process enough information and the AI is able to 'handle' all the information, it might be better to consider as many factors as possible in certain situations. The more factors and information humans and also AI consider at the same time, the closer they can come to making the perfect decision.

10.3. Reasons to Refrain from the Mens Rea Requirement

Similar to German law the mens rea requirement is abandoned for the reason that the protection of the public requires a high standard of care on the part of the persons performing risky activities. Such persons – especially if AI were to achieve personhood – have to take proper care. Without liability for, e.g. inadvertence, the defendant might be able to assign the potential costs of his behaviour to the society. At the same time, he will have no stimulus to at least reduce these risks. Threatening persons with criminal liability provides them a motive to apply measures while performing risky activities (risk-generating activities), which they probably would not have taken, to ensure safety and to eliminate unforeseen accidents or faults.[462] Here the question arises whether any

[460] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems.* Basel, Springer Cham, 2015, p. 99.
[461] Ibid., p. 100.
[462] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 169.

activities of robots may be seen as 'risky activities' solely based on the fact that the robot turned itself on, making it liable for its deeds, based on strict liability. Of course, this scenario is only applicable to strong AI. However, this this approach might be the key to applying strict liability by default to AI.

Further, such activities, controlled by regulatory offences, are based on the commercial behaviour of companies. Strict liability appears to be an exceptionally useful tool in cases where the offence is aimed at controlling the activities of companies. The reasons are mostly the fact that the conviction of the corporation does not usually involve the same standard/degree of stigma or the practical implications. Furthermore, the activities of corporations are usually on a larger scale than activities performed by individuals, which leads to the creation of greater levels of threat to society.[463] This might especially be true when AI is globally networked, hence machines would not necessarily be working on a fully individual basis, which may lead to some certain decisions having a worldwide influence.

However, the most important reason is that the proof of mens rea is particularly challenging in the context of corporate entities, primarily due to the absence of a specific individual who can be identified as the agent responsible for the corporation's actions. Strict liability is far easier to apply to companies, as it can be imposed without reference to the mental state of the defendant.[464] Comparable to companies, where there is no mind involved in the traditional sense, the application of strict liability should be the only suitable way to circumvent the problem of finding mens rea. The assignment of guilt is especially difficult within the use of autonomous machines. Once such a machine is started with a certain objective, its actions are to be understood as a diffusion of responsibility. As long as there is no evident mistake, clearly made by one party, criminal liability is not allowed to be imposed upon any one of them.[465]

---

[463] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 170.
[464] Ibid., p. 169.
[465] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 28.

Forbidding the use of autonomous machines in certain spheres might appear as an effective measure to avoid such complexities. However, banning autonomous machines constitutes a heavy restriction on the freedom of research and the general freedom to act of the involved parties. Further, such a solution would deprive society of the benefits that may be derived from the usage of machine decision-making.[466]

10.4. Strict liability and AI

10.4.1. Basics

Many offences against the criminal law do not involve a kind of public condemnation which is implicit in a conviction for homicide. Serious criminal offences, especially in cases where significant stigma is affiliated to the conviction, must stipulate a sort of mens rea. This form of mens rea should be proved before the conviction of the defendant to avoid criminalising a defendant who is not at fault of the occurred harm.[467] However, there are many possible offences where the mens rea element is inexistent. These offences lead to the imposition of strict liability. In practice it means that the prosecution requires actus reus to be proven, while the element of mens rea is absent. Thus, even if the defendant was not at fault, that person will be prosecuted nevertheless. Driving a car without a valid licence, for example, constitutes a strict liability offence.[468]

Such offences are nothing new to law, in fact this sort of liability was invented in the Roman empire where the owners of slaves were held responsible for the deeds of their slaves.[469] According to this, liability is linked to the conducting agent without the need to prove the fault. Therefore, liability for violations is imposed on the person who caused it, regardless of the fact that the responsible agent conducted the breach of behavioural norm with intention or

[466] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 28.
[467] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 157.
[468] Ibid., p. 159.
[469] Ibid., p. 157.

within a certain mental state.[470] This approach comes very close to the owner's liability for animals, and is therefore often applied to certain risk groups where the risk of any harm is directly attached to the foreseeability of their behaviour. In such cases, the person held liable is the person considered responsible for watching the animal, since they are the ones in the optimal position to take countermeasures in order to at least reduce the risk of harm.[471]

Before applying strict liability on AI, the question needs to be clarified as to which state of mens rea is required to fulfil the offence. The above-mentioned analysis shows the required mental elements of e.g. murder, such as intention, or at least foreseeability of the consequences from the conduct.

As stated above, mens rea means intention, recklessness, knowledge (about the possible outcome), foreseeability or belief (as appropriate).[472] This makes cases where the activity of the defendant is not necessarily unlawful, but may be an offence when the perpetrator carried out this action in a specific manner, more complicated. For instance, it is a legitimate activity to drive a vehicle. However, driving an uninsured car constitutes a strict liability offence. There are several relevant factors used to assess such offences. Strict liability is more likely applied in cases where the specific activity needs to be done by someone with specialist skills. The probability is higher that the regulation of the activity involves strict liability if the activity requires specialist skills. Strict liability makes the prosecution much simpler in cases where it is expected that the defendant is aware of the error and the possibilities to prevent it. In circumstances like this, it is irrational to assume that the prosecutor can adopt sufficient information of the defendant's activity to prove mens rea.[473]

UK case law for example has several important examples which show the reason and especially the importance of a liability regime, which at least partially exclude the existence of mens rea as a requirement for punishability. The

[470] Yeung, K. *Responsibility and AI: Council of Europe Study,* DGI(2019)05, Council of Europe, 2019, p. 60.
[471] Ibid., p. 61.
[472] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 161.
[473] Ibid., p. 162.

Alphacell v. Woodward (1972) judgment brought a highly sufficient reasoning why mens rea should not always be a requirement: "If [...] no conviction could be obtained under the 1951 Act unless the prosecution could discharge the often impossible onus of proving that the pollution was caused intentionally or negligently, a great deal of pollution would go unpunished and undeterred to the relief of many riparian factory owners."[474] Regulating a voluntary activity of a specialist does not result in convicting citizens for doing ordinary things without being aware of the fact that the respective action is in danger of becoming a criminal act. Thus, considerations about fair warning become less relevant if the defendant is said to have at least assumed the risks of liability, just by the fact that he voluntarily brought himself into the specific sphere of operation of a regulatory law.[475] The Hobbs v. Winchester Corporation (1910) judgment reinforces this reasoning. If someone chooses to engage in a business selling products which may be dangerous to health or, in some cases, even deadly, this person has to take that risk. Further, saying that the person could not have discovered the danger unless he had an analyst on the premises, is not considered a sufficient excuse for someone who chose to engage in such type of business.[476] The court made a similar point in the Sweet v. Parsley (1970) judgment: "Where penal provisions are of general application to the conduct of ordinary citizens in the course of their everyday life, the presumption is that the standard of care required of them in informing themselves of facts which would make their conduct unlawful, is that of the familiar common law duty of care. But where the subject-matter of a statute is the regulation of a particular activity involving potential danger to public health, safety or morals, in which citizens have a choice whether they participate or not, the court may feel driven to infer an intention of Parliament to impose, by penal sanctions, a higher duty of care on those who choose to participate […]"[477] However, mens rea is the standard property of an offence, its implication is unambiguously offset by other factors. As a result, in favour of strict liability, the presumption of mens rea is to be

[474] Alphacell Ltd. vs Woodward [1972] AC 824.
[475] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 163.
[476] Hobbs v. Winchester Corporation [1910] AC 132, 163.
[477] Sweet v. Parsley [1970] AC 132, 163.

suspended in cases where it would be unreasonable to understand the offence as such where mens rea is required.[478]

For the given question it is important to note that the UK courts are rather willing to apply strict criminal liability in cases where the statute regulates certain activities of specialists. Such laws are very often found within regulations such as the quality of food products, the distribution of drugs and safety on the road,[479] which is of most interest here.

10.4.2. Arguments against Strict Liability

As analysed above, there are many reasons to apply strict liability to AI-systems. However, equally, there are also reasons that oppose strict liability in the cases of robots. The major argument against strict liability, or to always require mens rea, is that the above-mentioned reasons do not justify why criminal rather than civil sanctions are to be used in order to adjust the activity in question. Even if natural or legal entities are unlikely to be able to claim compensation for damages, and especially so not from AI-systems, as do not possess the financial means, the possibility might be open to the government, after certain legislative restructuring, to claim charges from offenders through an administrative system of laws applying certain standards of strict liability without criminal convictions and its accompanying connotations.[480]

As shown above, the strict liability-approach might lead to a suitable solution for imposing criminal liability on AI. However, there are moral concerns in doing so. The question is whether it is morally acceptable to impose strict liability despite the absence of mens rea. Critics will likely say that with the absence of mens rea, strict liability will allow the conviction of blameless persons[481], or in this case machines. However, to follow this line of criticism would also make imposing strict liability to human beings questionable. The question whether doing the same with machines is even more complicated to justify. Nonetheless, keeping with English law as an example, one problem

[478] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 161.
[479] Ibid., p. 163.
[480] Ibid., p. 170.
[481] Ibid., p. 168.

appears to be that the English courts see strict liability as the only available alternative to the mens rea requirement, that is, of intention or recklessness.[482]

10.5. Networked AI-Machines

Determining the criminal liability within cases of networked autonomous machines is even more complicated. When such machines are generally accepted as being useful, society has to accept the shift of responsibility to these machines as well. Society will no longer be permitted to hold an individual as criminally liable for his actions. The liability of persons who produce, sell or use such machines need to be, in the spheres where such machines are generally allowed, of a shorter range compared to cases where the persons acted themselves. This applies especially to individuum-oriented criminal liability. It would be unjust to pass on the liability to an individual. However, this does not mean that liability for negligence should be negated. The more dangerous the respective application of autonomous machines is in a specific sphere, and the more it is used only for the benefit of the user, the more uncontrollable it is for the user and for third persons, and hence the more it seems plausible to apply liability due to negligence. This is despite the lack of the machine's foreseeability of the specific misdemeanour and damage. Such a case can occur within the application of armed machines which are exclusively made to harm, or even kill human beings. The fact that these machines may make a wrong decision is clear to every user. When such machines are used for private purposes exclusively, liability for negligence for such wrong decisions is applicable.[483] This applies especially to cases where such machines are connected to other machines, or to the internet given the due possibility that others may obtain access and share faulty information, which in turn leads to wrongful behaviour. However, exceptions should be made for machines designed to benefit all of society which accidentally cause any type of harm. If the social advantage requires the

[482] Simester, A./Sullivan, G. R. *Criminal Law: Theory and Doctrine.* In. *Juridical studies.* vol. 1, London, Bloomsbury Academic, 2000, p. 168.
[483] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 20.

autonomation, or networking of a specific system, one individual cannot be held criminally liable for that specific machine.[484]

Using machines which are capable of decision-making without any intervention of human beings bears significant risks of damaging the legally protected rights of the users and third parties per se. Firstly, machine-made decisions currently have a considerably higher error rate.[485] Secondly, machines lack the properties which play an important role within human interactions such as empathy and creativity.[486] This may lead to the fact that some machine-made decisions may seem inadequate or faulty from a human perspective.[487] The latter aspect is narrowly connected to the responsibility gap, which is heavily discussed from the ethical as well as legal perspective.[488] Human beings who evaluate, or are even affected by the decisions of machines are not only confronted with a lack of empathy, but further with the fact that the decision is being made without human interference and thus no human having to take the responsibility for its actions. However, it is the very purpose of such systems to assist human beings within certain decisions, such as day-to-day, or even war situations. The assistance is only noticeable if human beings cannot be made criminally liable for wrong decisions, or for decisions which are incomprehensible to human minds. Further, as previously examined, it is probably unjustifiable to blame someone alone for the faulty decision of a machine. Making the user extensively liable seems to be wrong, given the fact that the faulty decision may be based on numerous different reasons, such as inadequate programming, material defects and many other failures which cannot

---

[484] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 31.

[485] Ibid., p. 14; for examples of such wrong decisions see Müller-Hengstenberg, C. D./Kirn, S. *Intelligente (Software-) Agenten: Eine neue Herausforderung unseres Rechtssystems Rechtliche Konsequenzen der "Verselbstständigung" technischer Systeme.* In. *Multimedia und Recht.* 17(5), 2014, pp. 309 f.

[486] Becker, M. *Autonome Waffen – Wie Roboter den Krieg menschlicher machen sollen.* 2014, https://www.spiegel.de/wissenschaft/technik/autonome-waffen-roboter-sollen-krieg-ethischer-machen-a-977614.html.

[487] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 14.

[488] Matthias, A. *Automaten als Träger von Rechten*. 2nd ed. Berlin. Logos Verlag Berlin, 2010, pp. 13 ff.

be prevented by the user.[489] Further, autonomous machines are, or will be, that complex that the specific reason which leads to the error will most likely be impossible to detect.[490] Additionally, many autonomous machines will probably need extensive 'teaching' either from their user or another person so that they can participate in activities safely. For this reason alone, it would make no sense to make the programmer, producer or seller extensively liable for its deeds. This is comparable to the education of children. They need intensive teaching and education for (at least) eighteen years. Within this period of time, the parents are primarily responsible for the actions of their children. However, after eighteen years, when the person is considered an adult, it is unjustifiable to continue to blame the parents if the adult 'child' then commits a crime. In this regard, if there is any damage caused by a robot while it is still learning, its owner or user shall be liable for its deeds. However, there might by space for variations, depending on the whether the user is a professional or not.[491]

As set out above, the legal difficulties gain in complexity when the autonomous systems are networked. Such networks may have errors themselves, such as unstable connections, or they may be subject to hacking attacks or infested with the wrong information which the systems might use. Wrongful decisions may spread throughout the network and thus the autonomous machines, leading to a chain reaction.[492] The networked cooperation between a large number of machines, which make their own decisions, learned from their own experiences, following their own respective targets, increases the risk of unpredictability, and especially the uncontrollability of their decisions. Ultimately, the potential for conflicts lies in the fact that the problems of the internet interplay with the problems of the machine's autonomy. This sets up a

[489] Beck, S. *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In. Beck, S./Meier, B.-D./ Momsen, C. *Cybercrime und Cyberinvestigations.* 1st ed., vol. 6, Baden-Baden, Nomos, 2015, p. 14.
[490] Ibid., p. 15.
[491] Directorate-General for Internal Policies, Citizens' Rights and Constitutional Affairs. *European Civil Law Rules in Robotics. European Union,* 2016, p. 17.
[492] Gruber, M. In: Gruber, M. et al. *Was spricht gegen Maschinenrechte?* In. *Autonome Automaten - Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft.* 1st ed. Berlin, trafo Wissenschaftsverlag, 2014, p. 195.

place for legal spaces of absent liability, developing significant legal risks and dangers for society.[493]

Furthermore, the European parliament is considering designing future legislation based on an in-depth evaluation to decide between the application of strict liability or a risk-management approach. The Parliament notes that strict liability solely requires proof that the damage actually occurred together with the causal link between the harmful functioning of the machine and the damage that the injured party suffers. Further, the Parliament states that the risk management approach will not focus on the person who acted negligently and make him liable for the damaging act but rather the person who is, under certain circumstances, able to minimise the risks and to deal with the negative effects.[494]

When the ultimately responsible party is identified, their liability should be proportional to the actual extent of the commands given to the robots, as well as proportional to the degree of the robot's autonomy. The greater their learning ability, or autonomy is, and the longer the 'training' or 'education' of the robot takes, the greater the responsibility of their respective 'trainer' should be. At the same time, within the determination of the liable party, the abilities taught and given to the robot are not to be confused with the abilities which are to be fully attributed to its learning capabilities.[495] It should, however, be kept in mind that this concept is being developed for weak AI.

After the extensive analysis concerning the criminal liability of AI, the question arises whether it makes sense to create one universally applicable liability regime for AI. As mentioned above, there are two major types of AI: strong and weak AI. The differences between these two categories are significant, to put it mildly. In this regard, can it be the right way to develop an overreaching liability system for all types of AI? Formulating a general liability regime for AI does not make sense because the differences in autonomy and

[493] Beck, S. *Roboter und Cyborgs – Erobern sie unsere Welt?* In. Beck, S. *Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs. Robotik und Recht.* 1st ed. vol. 1, Baden-Baden, Nomos, 2012, p. 16.
[494] European Parliament. *Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.* P8_TA(2017)0051, recital 53 f.
[495] Ibid., recital 56.

intelligence are, as aforementioned, too wide.[496] It would be a better idea to make a case by case decision regarding the respective choice of liability regime, based on the specific robot category.[497] This means a technology-based approach, which considers liability relevant factors such as learning- and decision-making capabilities would be the preferred solution.[498]

The Member of the EU-Parliament, Mady Delvaux, states in an interview[499] concerning the resolution[500] of the EU-Parliament that the strict liability of the producer for its robots' deeds shall be examined. Delvaux adds that producers are those in the best position to minimise the risks. However, this solution is too cursory. Especially for robots with learning capabilities, the influence of their holder on their respective learning process is significant. As mentioned above, the holders socialise their robots by using them in certain specific environments.[501] It is the holder who chooses a suited robot for the specific task, who takes safety measures and who profits from the usage.[502] This solution is to be preferred, given that passing on the liability to the producer would run counter to the promotion of innovation, which is a goal of the aforementioned EU resolution.[503]

10.6. Applicability of Defences to AI-Systems

Machines equipped with strong AI might not plead the absence of mens rea or fault if charged with a strict liability offence, as the presence of mens rea is

---

[496] Lohmann, M. F. *Ein europäisches Roboterrecht - überfällig oder überflüssig?* In. *Zeitschrift für Rechtspolitik.* 06/2016, p. 169.
[497] Bertolini, A. *RoboLaw: Why and how to regulate robotics.* 2014, https://robohub.org/robolaw-why-and-how-to-regulate-robotics/
[498] Lohmann, M. F. *Ein europäisches Roboterrecht - überfällig oder überflüssig?* In. *Zeitschrift für Rechtspolitik.* 06/2016, p. 169.
[499] https://www.europarl.europa.eu/news/en/headlines/economy/20170109STO57505/rise-of-the-robots-mady-delvaux-on-why-their-use-should-be-regulated.
[500] European Parliament. *Report with recommendations to the Commission on Civil Law Rules on Robotics.* (2015/2103(INL)).
[501] Müller, M. F. *Roboter und Recht – Eine Einführung.* In. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle*. 05/2014, pp. 598 ff.
[502] Hanisch, J. In. *Zivilrechtliche Haftungskonzepte für Roboter.* Hilgendorf, E./Günther, J-P. *Robotik und Gesetzgebung – Beiträge der Tagung vom 7. Bis 9. Mai 2012 in Bielefeld*. *Robotik und Recht.* 1st ed. vol. 2, Baden-Baden, Nomos. 2013, p. 114.
[503] Lohmann, M. F. *Ein europäisches Roboterrecht - überfällig oder überflüssig?* In. *Zeitschrift für Rechtspolitik.* 06/2016, p. 169.

not required in such cases. However, there are a few available defences, such as duress, necessity, or self-defence, which need to be examined in terms of AI.

In cases of duress it must be considered that it probably requires AI which is able to sense, or to feel stimulus such as pain or distress. A robot which does not possess the ability to feel any negative emotions or even pain is hard to imagine in a situation of duress. Such a system is not naturally eager to survive or to extricate itself from dangerous or uncomfortable situations, as especially the latter situations do not exist for such machines. For this reason, cases of duress shall be excluded.

Necessity, however might be a relevant and applicable defence to AI. To revisit the example of a self-driving car facing a dilemma situation where it must choose to sacrifice one person to save another, the AI cannot simply do nothing. Even no reaction, aside from braking, is still a form of reaction. Therefore, the AI must make a decision and act accordingly. Here the question arises as to why the machine (or any other party behind the machine) should be held criminally liable. It may be a fact that strict liability often requires the action of a specialist, which may be a matter of fact in cases of AI. It can be argued that most AI systems will be made for one specific task so as to fulfil this at the highest possible level. Thus, AI systems are definitely 'specialists' in their given field. Still, it is not justified to blame AI, or whoever is involved with this technology if it gets into situations where there is no possible solution without harming or even killing a person, when other parties are at fault, since human beings are not held criminally liable in such situations as well. Considering such situations, strict liability of AI, especially in cases of autonomous driving needs to be carefully adjusted in terms of necessity.

Also, self-defence may be questionable but in certain situations acceptable. Examples may be the defence of the machine's owner or property. Even though self-defence may be close to situations of duress, it is certainly thinkable that at least some AI systems will be programmed or train themselves to have some kind of a 'survival instinct'.

10.7. Vicarious Liability

The justification of the existence of vicarious liability in criminal law is to create the possibility of holding a person liable as the principle offender. That means that the actus reus of a crime is physically not committed by the perpetrator but by someone else. Contrary to the above-mentioned liability-constellation, in cases of vicarious liability the person who physically performed the actus reus is not innocent, at least not entirely, and may be liable for the offence as well. Here, the legal emphasis lies on the relationship between the defendant and the person or entity physically performing the actus reus, and in some cases based on that relationship it attributes the acts of the deeds of the 'performer' to the perpetrator. This constellation is very often found within employee and employer relations and is similar to the principal and agent approach from the civil law.[504] In current cases, where there are hardly, if any robots involved, this form of liability is very rare and thus an exception. Generally, criminal liability is imposed on the basis of a person's acts rather than the acts of another. That is one significant difference to the approach of civil law (e.g. torts law) where vicarious liability is exceedingly common.[505] Applying vicarious liability to AI, there are most likely two possible scenarios: the first of these would refer to cases where the perpetrator uses a robot with weak AI to, for example, kill somebody. In such cases the solution is easy. Since weak AI is not 'intelligent', it is not aware of its actions, and is mostly dependent on instructions of its owner, since it is remote controlled, either by direct control of the user, or its actions are predetermined by programming. The second scenario would be where the perpetrator uses a strong AI machine to perform the wrongful act. In this case, the machine is indeed 'aware' of what it is doing – or it should at least have the information – that committing murder is against the law. Here the machine makes the decision to act against the law, rather than acting lawfully. Then, both parties should be liable in the same way as it would be the case if two human beings were involved in the case. A third variation would be where strong AI uses another AI to perform the murder. As already examined, there is no way to punish a human being for the machine's act. Thus, the strong AI as the

[504] Heaton, R. *Criminal Law Textbook.* 2nd ed. London. Oxford University Press, 2006, p. 459.
[505] Ibid.

perpetrator must be punished here, since it had a certain type of awareness or, at the very least, it was able to foresee the consequence of its action.

10.8. Corporate Liability

As seen above, corporate criminal liability, or a modified type of corporate criminal liability may be a solution to answer the question of how the law should treat the actus reus of an agent where the existence of 'true' mens rea cannot be proven, or is entirely inexistent.

Imposing criminal liability on companies is a comparatively new development. In general, companies may be criminally liable to the same extent as individuals. Legally, corporations are considered a different person and are separated from the individual members which constitute the organisation. However, the character of some offences precludes their commission by corporations, such as rape. Further, there is a problem in convicting companies for offences for which arrest is the only eligible punishment, which is also a significant problem as concerns AI. As mentioned above, a highly common example for corporate liability is the employer's liability for the actions of their employees. The practical effect of such cases is that the company is vicariously liable for all acts performed by the employees within the scope of their employment.[506] It might appear strange to hold companies criminally liable to just the same extent as human beings, since companies cannot carry out any actions without the agency of human beings. Neither can companies have a state of mind independent of human beings within the company. A fictional concept has been created by the (English) judicial system, whereby the actions and mental state of specific high-ranking executives or officials are considered to be the actions and mental state of their respective corporations.[507]

Concerning AI, it is highly important to mention that some legal systems, for example the English, can, under certain circumstances, hold companies criminally liable as accomplices for offences such as bigamy or even rape, even though for the latter, it is much harder to hold it liable as an accomplice.[508] The

[506] Heaton, R. *Criminal Law Textbook.* 2nd ed. London. Oxford University Press, 2006, p. 465.
[507] Ibid., p. 466.
[508] Ibid., p. 469.

UK Court of Appeal even accepted that corporations can be convicted of manslaughter.[509]

However, the personal responsibility of directors and executives may still come within corporate liability.

In cases where a company is liable, the person behind or who is connected to the action of the company will be liable as perpetrator. Other officials may be liable in the sense of a perpetrator or accessories as well. Of course, executives performing or authorising the acts in the name of the company have the possibility to 'hide' behind the company's criminal liability. The defendant (if it is a human being) has to possess mens rea for the respective crime. Further, mens rea is always required for accessories. However, personal liability is common in some cases when an offence is committed by a company, or any other executive officers. These persons will be held guilty in cases where the respective act was performed with the person's consent or connivance, or even negligence, if it is attributable to that person.[510] Consent refers to the fact that the defendant possesses knowledge of the events and agrees to them, while connivance connotes wilful blindness concerning the commission of the offence.

[509] Attorney-General's Reference No. 2 of 1999 [2000] 2 Cr App R 207.
[510] Heaton, R. *Criminal Law Textbook.* 2nd ed. London. Oxford University Press, 2006, p. 473.

Conclusion

Following a comprehensive review of relevant literature and legislation, including judicial rulings on criminal liability of AI, the question of whether amendments to criminal law are necessary in light of advances in AI technology remains largely unresolved. Nevertheless, the results of this research offer insight into the potential implications for the development of criminal law, including the direction of AI's criminal liability.

As this research demonstrates, there is a significant difference in the criminal liability of weak and strong AI. Therefore, it is imperative to consistently maintain a clear distinction between the two. Moreover, it is crucial to address the challenges surrounding the identification of the individual responsible for any malfunction in an AI system.

The research in this thesis shows that applying current criminal law to weak AI is unproblematic. Thus, one of the foremost challenges in the field of criminal liability of AI pertains to the identification of the individual who is culpable for any malfunction that may occur in the system. It is widely acknowledged that manufacturers bear primary responsibility for ensuring the operational efficiency of their products. However, the complexity of assigning liability in criminal cases involving AI demands a more detailed investigation of the parties involved, such as the producer, the programmer, and the user.

It is not straightforward to simply assign liability to the producer in situations where, for example, an autonomous vehicle fails to brake despite having sufficient time to respond appropriately, resulting in the death of a pedestrian. As such, the determination of the party accountable for an AI malfunction demands a thorough and meticulous investigation. In order to address this issue, the development of appropriate methodologies to accurately identify the responsible party becomes essential.

Once the responsible individual is identified, the implementation of criminal law, as alluded to earlier, presents no significant challenges. This issue presents a significant obstacle and necessitates the development of appropriate methodologies to accurately and definitively identify the responsible party.

Hence, the emphasis should be placed on determining culpability for a malfunctioning AI system, as opposed to contemplating the requisite legal alterations necessary to impose criminal liability on an individual.

This research illustrates the broad range of potential liability scenarios for weak AI that are currently well-established and applicable to issues such as autonomous driving, as well as various other applications of weak AI. Moreover, the need for legal adjustments should focus on creating guidelines for competent courts to effectively handle cases involving AI-related accidents and determine appropriate criminal liability.

In the light of the manifold existing legal models and ideas to allocate criminal liability, it appears that the one correct legal treatment of criminal actions performed by either strong or weak AI does not exist. This is logical as the variety of possible criminal deeds is even more extensive. Thus, we cannot expect one existing legal model of liability to allocate and impose criminal liability in all the different kinds of thinkable constellations of criminal deeds, or the negative consequences that arise from the use of AI. Therefore, it is crucial for lawmakers and legal experts to continuously analyse and adapt legal frameworks to keep pace with technological advancements and ensure adequate liability allocation.

An issue arises when considering strong AI, which is defined as having intelligence at least equal to that of a human being. Just as human beings, strong AI is not fully predictable in its decision-making behavior, so that the predictability of their misconduct can be questioned for the producer. If criminal law were to hold the producer or any other party liable for the actions of strong AI, it is unlikely that anyone would develop such a highly intelligent machine capable of learning and adapting, due to their unpredictability. This can be illustrated through a thought experiment: when a child is born, parents have certain obligations, one of which is accepting liability for the child's actions until they reach a certain age. Would it be fair to develop criminal law in a way that makes parents liable for their offspring's deeds regardless of age? Such a law would likely have significant implications for population development, as parents would have to fear the actions of their children for the rest of their lives.

Accordingly, it would be equally unjust to hold the producers (or any other party involved) of strong AI accountable for its actions if it possesses the intelligence, autonomy, and hence unpredictability of an adult person. This highlights the delicate balance that needs to be struck between accountability and fostering innovation in the field of AI.

Currently, strong AI does not possess the prerequisite of personhood required for criminal liability. Thus, if such AI were to exist today, it would not be subject to criminal liability for its actions. But given the fact that AI exists and has been attributed personhood, the question arises as to whether and how criminal law would need to be adjusted accordingly. This raises important philosophical and legal debates surrounding the definition of personhood and the implications it carries for AI ethics and responsibility.

Initially, there exists the potential for approaching AI with a framework similar to that applied to corporations within the field of criminal law in several jurisdictions. Such an approach would require substantial amendments or augmentations to existing criminal legislation, or the creation of specialized criminal law (i.e., lex specialis) designed for robots, as corporate criminal law fails to encompass transgressions like manslaughter or homicide. This highlights the need for innovative legal solutions that can effectively address the unique challenges posed by AI in the context of criminal liability.

The second option is to subject AI to criminal law, similarly to human beings. This study presents several compelling arguments in support of this approach. With regard to the mental element required for holding humans criminally liable, there is no evidence that humans possess consciousness, and thus, the fulfillment of any mental element during the commission of a crime cannot be scientifically demonstrated. Nonetheless, criminal law has managed to bypass the need for actual proof of any mental element in humans. The mere fact that the offender must have been aware of the occurrence of the outcome of their actions, which means foreseeability was present, is sufficient for criminal law to presume the mental element or the intention. Further, the assertion that intention may be imputed based on knowledge that the act was wrongful, coupled with a deliberate decision to proceed with the act, is equally

demonstrable for artificial intelligence as it is for human agents. Moreover, AI acquires factual data through sensors, such as cameras and microphones, in a manner that is analogous to the way in which humans use sensory organs, like the eyes and ears, to gather factual data. In humans, all the factual data gathered through their sensory organs are transported to the brain, where the information is consolidated into a final internal image. In robots equipped with strong AI, all the factual data collected through their sensors are transported to a central processing unit, where the final image is generated from all the factual data. The abovementioned factors suggest that under the given circumstances, mens rea is very much possible to find behind the deeds of machines. These parallels in the acquisition and processing of factual data imply that AI is capable of exhibiting mens rea, further strengthening the argument for subjecting AI to criminal law. They further suggest that the differences between humans and AI are not as significant as one might expect, which also supports the argument for subjecting strong AI to existing criminal law and applying criminal law to AI in the same way as it is applied to humans.

Given the non-provability of consciousness and the resemblances in the acquisition and manipulation of factual data, coupled with creation of the final image, it becomes apparent that distinguishing between human individuals and strong AI with regards to criminal law and consequent culpability is unnecessary.

Nevertheless, considering the present state of technology and the velocity of advancement, it is conceivable that the criminal culpability of strong artificial intelligence will remain a thought experiment in the foreseeable future.

The title of this thesis suggests a comparison of the criminal liability of strong and weak AI. The results show that the criminal responsibility of weak and strong AI has hardly any similarities. However, the differences between weak and strong AI can be clearly demonstrated. Weak AI lacks cognition, intelligence, and awareness of its deeds. For these reasons, weak AI cannot be a subject of criminal law. Weak AI is always used as a "tool" to support humans in various tasks and even to perform certain tasks completely autonomously. For this reason, a human being will always be criminally responsible behind weak AI. This study shows a variety of possible legal bases on which a person can be

held criminally liable for a crime caused by a malfunction of weak AI. The similarities in liability are therefore not to be found within weak and strong AI, but between the producer (or the respective guilty party behind weak AI) and strong AI itself. This is because strong AI has cognition, intelligence, and awareness, at least to an extent that the strong AI has demonstrably made conscious decisions, which is a fundamental requirement of criminal law. Therefore, according to current criminal law doctrine, there would be no obstacle to proving mens rea, and strong AI would be responsible before criminal courts, in contrast to weak AI. Therefore, the initial impression that the question of the criminal responsibility of weak AI is much simpler to answer than that of strong AI is misleading.

This thesis shows, in essence, that due to the many different possible persons behind weak AI, each of whom can be held criminally liable on a multitude of possible bases, the respective criminal court must work through a complex network of different possibilities in order to find a person guilty. This is not necessary for strong AI, as no one is likely to be behind strong AI. The results show that theoretically, it will not be entirely risk-free for companies to produce weak AI systems, whether for autonomous driving or other systems, as criminal liability could arise despite thorough testing and observation of the products already on the market. As shown above, not even suppliers would be safe, but individuals such as programmers would also not be fully safe from criminal liability for accidents caused by their AI systems. In this respect, there would indeed need to be certain adjustments in criminal law, but not legislative ones, rather certain guidelines on how the respective competent court should deal with such cases.

List of Abbreviations

AI	Artificial Intelligence
BGB	Bürgerliches Gesetzbuch (German Civil Code)
BGE	Swiss Federal Supreme Court
BGH	German Federal Court of Justice
BGHSt	Entscheidungen des Bundesgerichtshofes in Strafsachen (Germany)
BGHZ	Entscheidungen des Bundesgerichtshofes in Zivilsachen (Germany)
BVerfG	German Federal Constitutional Court
GSSt	Großer Senat für Strafsachen (Germany)
NJW	Neue Juristische Wochenschrift
StGB	Strafgesetzbuch (German/Swiss Criminal Code)
IFR	International Federation of Robotics
StR	Strafverfahren
PECL	Principles of European Contract Law
Vor.	Vorbemerkung
LIDAR	Light Detection and Ranging
NTSB	National Transportation Safety Board
GM	General Motors
BvR	Aktenzeichen einer Verfassungsbeschwerde zum Bundesverfassungsgericht

Bibliography

Antonio, Ianni/Monterossi, Michael William (2017). Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability. *Griffith Law Review,* 26(4), 563–592. https://doi.org/10.1080/10383441.2017.1558611.

Ashworth, Andrew (2006). *Principles of Criminal Law* (5th ed.). Oxford: Oxford University Press. ISBN: 978-0-1992-8114-5

Ashworth, Andrew/Horder, Jeremy (2013). *Principles of Criminal Law* (7th ed.). Oxford: Oxford University Press, ISBN 978-0-19-967268-4

Bartsch, Michael (2000). Computerviren und Produkthaftung. *Computer und Recht (CR)* 11/2000, 721–725

Beck, Susanne (2009). Grundlegende Fragen zum rechtlichen Umgang mit Robotik. *Juristische Rundschau (JR)*, 6/2009, 225–230. https://doi.org/10.1515/JURU.2009.225

Beck, Susanne (2012). *Roboter und Cyborgs – Erobern sie unsere Welt?* In Beck, Susanne (ed.), Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs. Robotik und Recht, (1st ed.), (vol. 1), pp. 9–22. Baden-Baden: Nomos 2012. doi.org/10.5771/9783845237527, ISBN: 978-3-8452-3752-7

Beck, Susanne (2015). *Google Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht?* In Beck, Susanne/Meier, Bernd-Dieter / Momsen, Carsten (eds.), Cybercrime und Cyberinvestigations, Robotik und Recht, (1st ed.), (vol. 6), Baden-Baden: Nomos. doi.org/10.5771/9783845266121. ISBN: 978-3-8487-2453-6

Beck, Susanne (2016). Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood. *Robotics and Autonomous Systems.* Volume 86/2016, 138-143. https://doi.org/10.1016/j.robot.2016.08.028

Becker, Markus (26 June 2014). *Autonome Waffen – Wie Roboter den Krieg menschlicher machen sollen.* Accessed: 05/03/2023,

https://www.spiegel.de/wissenschaft/technik/autonome-waffen-roboter-sollen-krieg-ethischer-machen-a-977614.html

Beer, Jenay M. / Fisk, Arthur D. / Rogers, Wendy A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, 3(2), 74-99. http://dx.doi.org/10.5898/JHRI.3.2.Beer

Bekey, George A. (2005). *Autonomous Robots – From Biological Inspiration to Implementation and Control*, MA, London: The MIT Press Cambridge.

Bekey, George A. (2012). Current Trends in Robotics: Technology and Ethics. In P Lin, K Abney, G Bekey (ed.), *Robot Ethics - The Ethical and Social Implications of Robotics,* Cambridge, Massachusetts: The MIT Press

Berger, Max B. (2016). Autonomes Fahren – Gedanken und Lösungsvorschläge. *Zeitschrift für Haftung und Versicherung HAVE,* 03/2016, 289 – 295.

Berster, Lars (2012). "Amoklauf von Winnenden" – zur Fahrlässigkeitshaftung neben der volldeliktischen Vorsatztat eines anderen. Note to the decision of the German Federal Court of Justice (BGH) case 22.3.2012 – 1 StR 359/11. *Zeitschrift für Internationale Strafrechtsdogmatik (ZIS),* 12/2012, 623–627

Bertolini, Andrea (29 October 2014). *RoboLaw: Why and how to regulate robotics.* Robohub. Accessed: 04.02.2023, from https://robohub.org/robolaw-why-and-how-to-regulate-robotics/

Boscarato, C. (2011). Who is responsible for a robot's actions? in van der Berg, B./Klaming, L., Technologies on the stand: Legal and ethical questions in neuroscience and robotics, Wolfpublisher, 383 – 402 *(18) Robots, market and civil liability: A European perspective*. Accessed: 06/05/2023, from: https://www.researchgate.net/publication/261351007_Robots_market_and_civil_liability_A_European_perspective [accessed May 06 2023].

Brenner, Walter/Zarnekow, Rüdiger/Wittig, Hartmut (1998). *Intelligente Softwareagenten – Grundlagen und Anwendungen.* Berlin: Springer. ISBN 978-3-662-12104-7

Bridgeman, Bruce (1980). Brains + programs = minds. *Behavioral and Brain Sciences,* 3(3), 427–428. DOI: 10.1017/s0140525x00005781

Brox, Hans/Walker, Wolf-Dietrich (2010). *Allgemeiner Teil des BGB* (34th ed.). Munich: Verlag Franz Vahlen. 978-3-8006-4141-3

Brox, Hans/Walker, Wolf-Dietrich (2013). *Allgemeiner Teil des BGB* (37th ed.). Munich: Verlag Franz Vahlen. ISBN 978-3-8006-3992-2

Card, Richard (ed.) (2006). *Card Cross and Jones Criminal Law* (17th ed.). Oxford: Oxford University Press. ISBN: 978-0-19-928666-9

Chopra, Samir/White, Laurence F. (2011). *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press. DOI: 10.3998/mpub.356801

Christaller, Thomas/Decker, Michael/Gilsbach, M. Joachim/Hirzinger, Gerd/Lauterbach, Karl W./Schweighofer, Erich/Schweitzer, Gerhard/Sturma, Dieter (2001). *Robotik und menschliches Handeln.* In *Robotik. Wissenschaftsethik und Technikfolgenbeurteilung.* Vol. 14. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-56422-2_4

Clarkson Christopher M.V./Hill, Jonathan (1976). *The Conflict of Laws* (3rd ed.). Oxford, New York: Oxford University Press 2006.

Columbus, Louis (2017, July 9). *McKinsey's State of Machine Learning And AI.* Accessed: 11/02/2023. https://www.forbes.com/sites/louiscolumbus/2017/07/09/mckinseys-state-of-machine-learning-and-ai-2017/?sh=1de4387175b6

Cording, Clemens / Roth, Gerhard (2015). Zivilrechtliche Verantwortlichkeit und Neurobiologie – ein Widerspruch? *Neue Juristische Wochenschrift,* 68(1-2), 26 ff.

Cornelius, Kai (2002). Vertragsabschluss durch autonome elektronische Agenten. *Multimedia und Recht (MMR),* 5(6), 353–358

Darling, Kate (23 April 2012). *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects.* Robot Law, Calo, Froomkin, Kerr eds., Edward Elgar 2016, We Robot

Conference 2012, University of Miami. Accessed:11/02/2023 https://ssrn.com/abstract=2044797 / http://dx.doi.org/10.2139/ssrn.2044797

Deutscher Bundestag, (German Federal Parliament) (2018). *Autonomes und automatisiertes Fahren auf der Straße – rechtlicher Rahmen*. Wissenschaftliche Dienste (WD) 3000 111/18

Directorate-General for Internal Policies, Citizens' Rights and Constitutional Affairs (2016). *European Civil Law Rules in Robotics.* European Union, ISBN: 978-92-846-0410-04

Donatsch, Andreas/Tag, Brigitte (2013). *Strafrecht I: Verbrechenslehre*. Zürich: Schulthess Juristische Medien. ISBN: 978-3-7255-6782-9

Ethik-Kommission (German Ethics Committee). (2017) *Automatisiertes und Vernetztes Fahren.* Report of 2017. Bundesministerium für Verkehr und digitale Infrastruktur. Accessed: 29/01/2023, from https://bmdv.bund.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile

European Parliament (12 January 2017). *Erste EU-weite "Robotergesetze": Interview mit Mady Delvaux* (Interview with Mady Delvaux). Accessed: 11/03/2023, from https://www.europarl.europa.eu/news/de/headlines/economy/20170109STO57505/erste-eu-weite-robotergesetze-interview-mit-mady-delvaux

European Parliament (12 January 2017). Rise of the robots: Mady Delvaux on why their use should be regulated. *News European Parliament*. Accessed: 23.04.2023, from https://www.europarl.europa.eu/news/en/headlines/economy/20170109STO57505/rise-of-the-robots-mady-delvaux-on-why-their-use-should-be-regulated

European Parliament (12 January 2017). *Rise of the robots: Mady Delvaux on why their use should be regulated*. Accessed: 22/01/2023 from https://www.europarl.europa.eu/news/en/headlines/economy/20170109STO57505/rise-of-the-robots-mady-delvaux-on-why-their-use-should-be-regulated

European Parliament (27 January 2017). Report with recommendations to the Commission on Civil Law Rules on Robotics, (2015/2103(INL))

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, P8_TA(2017)0051

European Parliament, Official Journal of the European Union, C 252/25, 18.7.2018, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.

Executive Office of the President National Science and Technology Council Committee on Technology *Preparing for the Future of Artificial Intelligence.* (2016). Accessed: 24/03/2023, from https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Feinberg, Joel (1987). *Harm to Self.* The Moral Limits of the Criminal Law (Vol. 3), Oxford: Oxford University Press, https://doi.org/10.1093/0195046641.001.0001.

Fellmann, Walter/von Büren-von Moos, Gabrielle (1993). *Grundriss der Produkthaftpflicht.* Bern: Stämpfli Verlag. ISBN: 978-3-72729-193-7

Fischer, Eugen/Vossenkuhl, Wilhelm (eds.) (2003). *Die Fragen der Philosophie: Eine Einführung in Disziplinen und Epochen.* Munich: C.H. Beck. ISBN: 978-340-649485-7

Fischer, Thomas (2016). *Strafgesetzbuch mit Nebengesetzen.* München: C.H. Beck. ISBN: 978-3-40668-260-5

Foerste, Ulrich/Graf von Westphalen, Friedrich (2012). *Produkthaftungshandbuch* (3rd ed.). Munich: C.H. Beck. ISBN: 978-3-406-60387-7

Freytag, Urs (2016). Sicherheitsrechtliche Aspekte der Robotik. *Sicherheit & Recht/Sécurité & Droit,* 02/2016, pp. 111 ff.

Frisch, Wolfgang (1983). *Vorsatz und Risiko – Grundfragen des Tatbestandsmäßigen Verhaltens und des Vorsatzes: Zugleich ein Beitrag zur*

Behandlung außertatbestandlicher Möglichkeitsvorstellungen. Köln, Berlin, Bonn, München: Heymann. ISBN: 978-3-452-19344-5

Frister, Helmut (2020). *Strafrecht Allgemeiner Teil* (9th ed.). Munich: C.H.Beck. ISBN: 978-3-406-75933-8. doi.org/10.17104/9783406763069

Gitter, Rotraud (2007). *Softwareagenten im elektronischen Geschäftsverkehr – Rechtliche Vorgaben und Gestaltungsvorschläge.* Baden-Baden: Nomos. ISBN 978-3-8329-3242-8

Gless, Sabine (2013). Strafrechtliche Produkthaftung. *Recht – Zeitschrift für juristische Weiterbildung und Praxis.* 2/2013, 54–64

Gless, Sabine (2016). *"Mein Auto fuhr zu schnell, nicht ich!" – Strafrechtliche Verantwortung für hochautomatisiertes Fahren*. In Gless, S./Seelmann, K (eds.), Intelligente Agenten und das Recht. Robotik und Recht, (1st ed.), (vol. 9), p. 225 ff. Baden-Baden: Nomos. ISBN 978-3-8452-8006-6

Gless, Sabine/Seelmann, Kurt (2016*). Intelligente Agenten und das Recht – Verantwortungszuschreibung in Antike und Moderne.* In Gless, S. / Seelmann, K (eds.), Intelligente Agenten und das Recht*,* Robotik und Recht, (1st ed.), (vol. 9), 11–20. Baden-Baden: Nomos. doi.org/10.5771/9783845280066. ISBN 978-3-8452-8006-6

Gless, Sabine/Weigend, Thomas (2014). Intelligente Agenten und das Strafrecht. *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW),* 126(3), 561-591

Görz, Günter/Nebel, Bernhard (2015). *Künstliche Intelligenz* (1st ed.). Frankfurt am Main: S. Fischer. ISBN 978-3-596-30136-2

Gramlich, Ludwig/Kröger, Detlef/Schreibauer, Marcus (2003). *Rechtshandbuch B2B Plattformen – Rahmenbedingungen elektronischer Marktplätze*. Munich: C.H. Beck. ISBN 978-3-406-49500-7

Gruber, Malte – Christian (2014). *Was spricht gegen Maschinenrechte?* In Gruber, Malte/Bung, Jochen/Ziemann, Sascha (ed.), Autonome Automaten – Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, (1st edition), Beiträge zur Rechts-, Gesellschafts- und Kulturkritik, (vol. 12). Berlin: trafo

Wissenschaftsverlag. ISBN: 9783864640452

Günther, Jan-Philip/Münch, Florian/Beck, Susanne/Löffler, Serverin/Leroux, Christophe / Labruto, Roberto (2012). Issues of privacy and electronic personhood in robotics. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication,* Paris, 2012. doi: 10.1109/ROMAN.2012.6343852.

Günther, Jens/Böglmüller, Matthias (2017). Künstliche Intelligenz in der Arbeitswelt*. Betriebs Berater (BB),* 2017, 53–58.

Haggard, Patrick/Lau, Hakwan (2013). What is volition? *Experimental Brain Research,* 229(3), 285-287. https://doi.org/10.1007/s00221-013-3582-5

Hallevy, Gabriel (2015). *Liability for Crimes Involving Artificial Intelligence Systems.* Basel: Springer Cham. https://doi.org/10.1007/978-3-319-10124-8

Hanisch, Jochen (2010). *Haftung für Automaten.* Internationale Göttinger Reihe (Vol. 19), Göttingen: J.-P. Cuvillier, ISBN: 978-3-8695-5319-1

Hanisch, Jochen (2013). *Zivilrechtliche Haftungskonzepte für Roboter.* In Hilgendorf, Eric/Günther, Jan-Philip (eds.). Robotik und Gesetzgebung – Beiträge der Tagung vom 7. Bis 9. Mai 2012 in Bielefeld, Robotik und Recht, (vol. 2), (1st ed.). Baden-Baden: Nomos. doi.org/10.5771/9783845242200 / ISBN: 978-3-8452-4220-0.

Harke, Jan Dirk (2016). *Sklavenhalterhaftung in Rom*. In Gless, S./Seelmann, K (eds.), Intelligente Agenten und das Recht. Robotik und Recht, (1st ed.), (vol. 9), p. 225 ff. Baden-Baden: Nomos. ISBN 978-3-8452-8006-6

Hart, H. L. A. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law* (2nd ed.). New York: Oxford University Press, ISBN: 978-019-953478-4.

Hawking, Stephen/Mlodinow, Leonard (2010). *The Grand Design.* New York: Random House Publishing Group. ISBN: 9780553907070

Heaton, Russel. 2006. *Criminal Law Textbook,* (2nd ed.). London: Oxford University Press. ISBN: 978-019-928705-5

Heckmann, Dirk (ed.) (2007). *juris PraxisKommentar Internetrecht* (1st ed.). Saarbrücken: juris, ISBN: 978-393-875606-5

Hefermehl, Wolfgang/Soergel, Hans-Theodor (1999). *Bürgerliches Gesetzbuch mit Einführungsgesetz und Nebengesetzen (BGB)* (vol. 2), (13th edition). Köln: Kohlhammer. ISBN 978-3-17-015793-4

Heintschel-Heinegg, Bernd von (ed.) (2016). *Beck'scher Online Kommentar StGB* (32nd ed.). Munich: C.H. Beck

Herberger, Maximilian (2018). "Künstliche Intelligenz" und Recht. *Neue Juristische Wochenschrift (NJW)*, 39/2018, 2825–2829.

Hilgendorf, Eric (2012). Können Roboter Schuldhaft handeln? In Beck, Susanne (ed.), Jenseits von Mensch und Maschine – Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs. Robotik und Recht, (1st ed.), (vol. 1), pp. 119-132. Baden-Baden: Nomos 2012. doi.org/10.5771/9783845237527, ISBN: 978-3-8452-3752-7

Hilgendorf, Eric (2015). Automatisiertes Fahren und Recht. *53. Deutscher Verkehrsgerichtstag 2015,* pp. 55–72. Köln: Luchterland, ISBN 978-3-472-08625-3

Hilgendorf, Eric (2018). *The Dilemma of autonomous driving: Reflections on the moral and legal treatment of autonomic collision avoidance systems.* Hilgendorf, Eric/Feldle, Jochen (eds.), Digitization and the Law, Robotik und Recht, (vol. 15), (1st ed.). Baden-Baden: Nomos. https://doi.org/10.5771/9783845289304-57/ ISBN: 978-3-8452-8930-4.

Hömig, Dieter/Wolff, Heinrich A. (eds.) (2022). *Grundgesetz für die Bundesrepublik Deutschland* (13th edition). Baden-Baden: Nomos. ISBN: 978-3-8487-7930-7

Hörnle, Tatjana (2016). Guilt and Choice in Criminal Law Theory. *Bergen Journal of Criminal Law and Criminal Justice,* 4(1), 1–24. https://doi.org/10.15845/bjclcj.v4i1.1023

Ianni, Antonio/Monterossi, Michael William (2017). Artificial autonomous agents and the question of electronic personhood: a path between subjectivity and liability. *Griffith Law Review,* 26(4), 563–592. https://doi.org/10.1080/10383441.2017.1558611

Jackson, Brandon W. (2019). Artificial intelligence and the fog of innovation: deep-dive on governance and the liability of autonomous systems. *Santa Clara High Technology Law Journal,* 35(4), 35–63.

Jänich, Volker M./Schrader, Paul T./Reck, Vivian (2015). Rechtsprobleme des Autonomen Fahrens. *Neue Zeitschrift für Verkehrsrecht (NZV),* 7/2015, 313 ff.

Janssens, Liisa (2018). A Prospect of the Future: How Autonomous Systems May Qualify as Legal Persons. In Janssens, L./Bayamlioglu, E./Baraliuc, I./Hildebrandt, M. (eds.), *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen,* pp. 116–121. Amsterdam University Press. https://doi.org/10.2307/j.ctvhrd092.24

Joecks, Wolfgang/Miebach, Klaus (eds.) (2017). *Münchener Kommentar zum StGB* (vol. 4), (3rd edition.) Munich: C.H.Beck. ISBN 978-3-406-68554-5

Joecks, Wolfgang/Miebach, Klaus (eds.) (2020). *Münchener Kommentar zum StGB* (vol. 4), (4th edition). Munich: C.H.Beck. ISBN 978-3-406-74601-7

John, Robert (2007). *Haftung für künstliche Intelligenz – Rechtliche Beurteilung des Einsatzes intelligenter Softwareagenten im E-Commerce*. Recht der Neuen Medien, (vol. 45). Hamburg: Dr. Kovac. ISBN 978-3-8300-3270-0

Kant, Immanuel (1781). *Kritik an der Vernunft.* Weischedel, W. (ed.) (1975). *Kant, Werke in Zehn Bänden* (vol. 6). Darmstadt: Wissenschaftliche Buchgesellschaft WBG.

Keil, Geert (2009). *Willensfreiheit und Determinismus* (2nd ed.). Ditzingen: Philipp Reclam jun. GmbH & Co. KG. ISBN: 978-3-15-019524-6.

Kindhäuser, Urs/Neumann, Ulfried/Paeffgen, Hans-Ullrich (ed.) (2013). *Strafgesetzbuch* (4th edition). Vol. 1. Baden-Baden: Nomos. ISBN: 978-3-8329-6661-4

Kirn, Stefan (1996). *Integration von Organisation und Informationssystem: Benötigen wir eine Re-Vitalisierung des maschinellen Aufgabenträgers?* Work report, Technische Universität Ilmenau.

Kirn, Stefan (2002). Kooperierende intelligente Softwareagenten. *Wirtschaftsinformatik,* 44(1), 53-63. https://doi.org/10.1007/BF03251465

Kirn, Stefan/Müller – Hengstenberg, Claus D. (2014). Intelligente (Software-) Agenten: Von der Automatisierung zur Autonomie? – Verselbständigung technischer Systeme. *MultiMedia und Recht (MMR),* 4/2014, 225 ff.

Koch, Frank (2005). *Internet-Recht* (2nd ed.). Berlin: De Gruyter Oldenbourg. ISBN: 978-3-486-57801-0

Kroll, Joshua A. (2015). *Accountable Algorithms* [doctoral dissertation, Princeton University]. Princeton University Doctoral Dissertations. http://arks.princeton.edu/ark:/88435/dsp014b29b837r

Kroll, Joshua A./Huey, Joanna/Barocas, Solon/Felten, Edward W./Reidenberg, Joel R. Robinson, David G./Harlan Yu (2017). Accountable Algorithms, *University of Pennsylvania Law Review* 165(3), 633.

Kuehn, Johannes/Haddadin, Sami (2017). An Artificial Robot Nervous System To Teach Robots How To Feel Pain And Reflexively React To Potentially Damaging Contacts. in *IEEE Robotics and Automation Letters*, (2)1, 72–79 doi: 10.1109/LRA.2016.2536360.

Küng, Hans (2008). *Anfang aller Dinge.* Munich: Piper Verlag. ISBN: 9783492251686

Kuri, J. (2016). *Tödlicher Tesla-Unfall: Autopilot hielt Lastwagen-Anhänger für hohes Schild*. Accessed: 08/01/2023

https://www.heise.de/newsticker/meldung/Toedlicher-Tesla-Unfall-Autopilot-hielt-Lastwagen-Anhaenger-fuer-hohes-Schild-3253449.htm

Lackner, Karl/Kühl, Kristian (2014). *Strafgesetzbuch* (28th ed.). Munich: C.H. Beck. ISBN: 978-3-406-70029-3.

Lackner, Karl/Kühl, Kristian (2018). *Strafgesetzbuch* (29th ed.). Munich: C.H. Beck. ISBN: 978-3-406-70029-3.

Legg, Shane/Hutter, Marcus (2007). Universal Intelligence: A Definition of Machine Intelligence. *Minds & Machines*, 17(4), 391-444

Lin, Patrick (2015). Why Ethics Matters for Autonomous Cars. In Maurer, Markus/Gerdes, J. Christian/Lenz, Barbara/Winner, Hermann (eds.), *Autonomes Fahren,* 69-85. Berlin Heidelberg: Springer Vieweg. https://doi.org/10.1007/978-3-662-45854-9_4

Locke, John (1690). *An Essay Concerning Human Understanding.* § XVII No. 26 331.

Lohmann, Melinda F. (2016). Ein europäisches Roboterrecht – überfällig oder überflüssig? *Zeitschrift für Rechtspolitik (ZRP),* 06/2016, 168 ff.

Lutz, Lennart S. (2015). Autonomes Fahren als rechtliche Herausforderung. *Neue Juristische Wochenschrift (NJW),* 3/2015, 119 ff.

Lycan, William G. (1990). *Mind and Cognition: A Reader.* Hoboken, New Jersey: Blackwell Publishers. ISBN: 978-0-63116-076-2

Markwalder Nora, / Simmler, Monika (2017). Roboterstrafrecht – Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle (AJP/PJA).* 02/2017, 171 ff.

Matthias, Andreas (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology.* 6(3), 175–183, https://doi.org/10.1007/s10676-004-3422-1

Matthias, Andreas (2010). *Automaten als Träger von Rechten* (2nd ed.). Berlin: Logos Verlag Berlin, ISBN 978-3-8325-1902-5

Mayer, Hellmuth (1967). *Strafrecht, allgemeiner Teil.* Stuttgart: Kohlhammer

McFarland, David/Bösser, Thomas (1993). *Intelligent Behaviour in Animals and Robots, MIT Press*. ISBN 978-0262132930

Müller – Hengstenberg, Claus D./Kirn, Stefan (2014). Intelligente (Software-) Agenten: Eine neue Herausforderung unseres Rechtssystems Rechtliche Konsequenzen der "Verselbstständigung" technischer Systeme. *Multimedia und Recht (MMR)*, 17(5), 307-313

Müller – Hengstenberg, Claus D./Kirn, Stefan (2016). *Rechtliche Risiken autonomer und vernetzter Systeme – Eine Herausforderung.* Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110431445

Müller, Melinda F. (2014). Roboter und Recht – Eine Einführung. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle (AJP/PJA).* 05/2014, 595 ff.

Murch, Richard/Johnson, Tony (2000). *Agententechnologie: Eine Einführung. Intelligente Softwareagenten auf Informationssuche im Internet* (1st ed.). Boston: Addison-Wesley. ISBN: 978-3-827-31652-3

National Transportation Safety Board, Preliminary Report Highway HWY18MH010. Accessed: 09.01.2023, from https://www.wsj.com/public/resources/documents/NTSBuber.pdf. Full report available at: https://www.wsj.com/public/resources/documents/NTSBuber.pdf. Last access: 06.05.2023.

Neuhäuser, Christian (2014). *Roboter und moralische Verantwortung.* In Hilgendorf. Erik (ed.). Robotik und Recht im Kontext von Recht und Moral, 1st ed. Vol. 3, pp. 269 – 286, Baden-Baden: Nomos, ISBN: 978-3-8487-1015-7, also available at: doi.org/10.5771/9783845252179

Niggli, Marcel Alexander/Wiprächtiger, Hans (eds.) (2013), *Basler Kommentar Strafrecht I* (3rd edition). Basel: Helbing & Lichtenhahn. ISBN: 978-3-7190-2980-7

Prittwitz, Cornelius (1993). *Strafrecht und Risiko – Untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft.* Juristische Abhandlungen (Vol. 22), (1st ed.). Frankfurt am Main: Vittorio Klostermann. ISBN 978-3-465-02587-0

Raz, Joseph (1988). *The Morality of Freedom*, Oxford: Oxford University Press. https://doi.org/10.1093/0198248075.001.0001.

Rengier, Rudolf (2019). *Strafrecht Allgemeiner Teil* (11th ed.). Munich: C.H. Beck. ISBN: 978-3-4067-3696-4

Russell, Stuart / Norvig, Peter (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Hoboken: Prentice Hall. ISBN 978-0-13-604259-4

Säcker, Franz Jürgen/ Rixecker, Roland/Oetker, Hartmut/Limpberger, Bettina (eds.) (2017). *Münchener Kommentar zum Bürgerlichen Gesetzbuch (*7th edition) (Vol. 6). München: C.H. Beck. ISBN: 978-3-40666-546-2

Scheutz, Matthias/Crowell, Charles R. (2007). *The Burden of Embodied Autonomy: Some Reflections on the Social and Ethical Implications of Autonomous Robots.* In Workshop on Roboethics at the International Conference on Robotics and Automation. 2007.

Schirrmacher, Frank (2013). *EGO-Das Spiel des Lebens.* Munich: Karl Blessing Verlag. ISBN 978-3-8966-7427-2

Schönke, Adolf/Schröder, Horst (eds.) (2010). *Strafgesetzbuch* (28th ed.). Munich: C.H. Beck, ISBN: 978-3-406-60404-1

Schönke, Adolf/Schröder, Horst (eds.) (2014). *Strafgesetzbuch* (29th ed.). Munich: C.H. Beck, ISBN: 978-3-406-652264

Schönke, Adolf/Schröder, Horst (eds.) (2019). *Strafgesetzbuch* (30th ed.). Munich: C.H. Beck, ISBN: 978 3 406 70383 6

Schweighofer, Erich/Menzel, Thomas/Kreuzberger, Hanna Maria (eds.) (2001). *Auf dem Weg zur ePerson: aktuelle Fragestellungen der Rechtsinformatik.*

Schweighofer, Erich/Lachmayer, Friedrlich (eds.) Schriftenreihe Rechtsinformatik (Vol. 3). Vienna: Verlag Österreich. ISBN: 978-3-7046-1719-4

Searle, John R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences,* 3(3), pp 417-424. https://doi.org/10.1017/S0140525X00005756.

Seelmann, Kurt (1994). *Privatrechtlich begründete Garantenpflichten?* Schmidt, Karsten (ed.), *Vielfalt des Rechts – Einheit der Rechtsordnung?* Hamburger Rechtsstudien, iss. 85, 1994, Berlin: Duncker und Humblot

Seelmann, Kurt (2007). Nichtstun als Straftat – Bedeutungszuwachs und Problempotential der unechten Unterlassungsdelikte. *Schweizerische Zeitschrift für Strafrecht (ZStrR),* 3/2007, 262 - 279

Seelmann, Kurt / Geth, Christopher (2016). *Strafrecht Allgemeiner Teil* (6th edition). Basel: Genf, München: Helbing & Lichtenhahn. ISBN: 978-3-7190-3556-3

Shahan, Zachary (30 June 2016). *1st Tesla Autopilot Fatality ... After 130 Million Miles (Updates)*. CleanTechnica. Last acces: 04.03.2023, from https://cleantechnica.com/2016/06/30/1st-tesla-autopilot-fatality-130-million-miles/

Shute, Stephen/Simester, Andrew (eds.) (2002). *Criminal Law Theory: Doctrines of the General Part.* Oxford Monographs on Criminal Law and Justice, pp. 171–206, Oxford: University Press. https://doi.org/10.1093/acprof:oso/9780199243495.003.0008

Simester, Andrew/Sullivan G. R. (2000). *Criminal Law: Theory and Doctrine.* Juridical studies (Vol. 1). London, Bloomsbury Academic, ISBN: 9781901362619

Simmler, Monika/Markwalder, Nora (2017). Roboter in der Verantwortung? – Zur Neuauflage der Debatte um, den funktionalen Schuldbegriff. *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW),* 129(1), 20–47. https://doi.org/10.1515/zstw-2017-0002

Söbbing, Thomas (2019). *Fundamentale Rechtsfragen künstlicher Intelligenz.* Frankfurt am Main: Fachmedien Recht und Wirtschaft. ISBN: 978-3-8005-1700-8

Solum, Lawrence B. (1992). *Legal Personhood for Artificial Intelligences.* North Carolina Law Review 70(4), pp. 1231 ff. Accessed: 11/02/2023 http://scholarship.law.unc.edu/nclr/vol70/iss4/4

Spindler, Gerald (2015). Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen. *Computer und Recht (CR)* 12/2015, 766-776. https://doi.org/10.9785/cr-2015-1205.

Spindler, Gerald/Schuster, Fabian (2019). *Recht der elektronischen Medien* (4th ed.). Munich: C.H. Beck. ISBN 978-3-406-73012-2.

Spitz, Phillippe (2001). *Strafrechtliche Produkthaftung – Übertragbarkeit zivilrechtlicher Betrachtungswesen? Unter besonderer Berücksichtigung der Organisationshaftung in Straf- und Zivilrecht.* In Basler Studien zur Rechtswissenschaft, (vol.12). [doctoral dissertation, University of Basel]. Basel: Basel, Helbing & Lichtenhahn.

Staudinger, Julius von (2022), *BGB Eckpfeiler des Zivilrechts – Kommentar zum Bürgerlichen Gesetzbuch und Nebengesetzen*. Berlin: ottoschmidt – De Gruyter. ISBN: 978-3-8059-1351-5

Stratenwerth, Günther (2011). *Schweizerisches Strafrecht, Allgemeiner Teil I: Die Straftat* (4th ed.). Bern: Stämpfli Verlag AG. ISBN: 978-3-7272-8667-4

Tafani, Daniela (2019). Dilemmata der Maschinen. Künstliche Intelligenz, Ethik und Recht. *Jahrbuch der Juristischen Zeitgeschichte*, 20(1), 302–322. https://doi.org/10.1515/jajuz-2019-0017

The Law Commission, Consultation Paper No. 177, A new Homicide Act for England and Wales? Accessed: 06.05.2023, from https://www.lawcom.gov.uk/app/uploads/2015/03/cp177_Murder_Manslaughter_and_Infanticide_consultation_overview_.pdf

Tzafestas, Spyros G (2016). *Roboethics. A Navigating Overview*, Berlin: Springer. 978-3-319-21714-7

Vladeck, David C. (2014). Machines without principals: Liability rules and artificial intelligence. *Washington law review*, 89(1), 117–150

Welzel, Hans (1961). *Fahrlässigkeit und Verkehrsdelikte: Zur Dogmatik der fahrlässigen Delikte.* Schriftenreihe Juristische Studiengesellschaft *Fahrlässigkeit und Verkehrsdelikte: Zur Dogmatik der fahrlässigen Delikte.* Heidelberg: C.F. Müller.

Welzel, Hans (1969). *Das deutsche Strafrecht* (11th edition.). Berlin: De Gruyter. ISBN: 978-3-11-000991-0

Wettig, Steffen (2010). *Vertragsschluss mittels elektronischer Agenten.* In Schriften zur Rechtswissenschaft, (vol. 142). [doctoral dissertation, University of Bayreuth]. Berlin: Wissenschaftlicher Verlag Berlin. ISBN: 978-3-86573-563-8

Wildhaber, Isabelle (2016). Die Roboter kommen – Konsequenzen für Arbeit und Arbeitsrecht. *Zeitschrift für Schweizerisches Recht (ZSR),* 135(4), 315–351.

Wildhaber, Isabelle (2017). Roboterrecht – Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse. *Aktuelle Juristische Praxis/Pratique Juridique Actuelle (AJP/PJA).* 2/2017, pp. 213 ff.

Williams, Glanville (1961). *Criminal Law: The General Part* (2nd ed.). London: Stevens & Sons Ltd.

Wohlers, Wolfgang (2016). Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen. *Basler juristische Mitteilungen,* 3/2016, 113–137

Wooldridge, Michael (2009). *An Introduction to MultiAgent Systems* (2nd edition). Hoboken: John Wiley & Sons. ISBN 9780470519462

Yeung, Karen. (2019). Responsibility and AI: Council of Europe Study DGI(2019)05. Council of Europe. Accessed: 04/02/2022, from https://rm.coe.int/responsability-and-ai-en/168097d9c5

Zarnekow, Rüdiger (1999). *Softwareagenten und elektronische Kaufprozesse*. Berlin: Springer. ISBN 978-3-322-95205-9

Zech, Herbert (2013). Gefährdungshaftung und neue Technologien. *JuristenZeitung (JZ)*, 68(1), 21–29.

Zurkinden, Nadine (2016). Strafrecht und selbstfahrende Autos – ein Beitrag zum erlaubten Risiko. *recht – Zeitschrift für juristische Weiterbildung,* 03/2016, 144–156

Table of Cases

Germany

Switzerland

United Kingdom

Canada

United States of Amerika

Abstract:

The rapid development of artificial intelligence (AI) has raised important legal and ethical questions regarding the potential criminal liability of AI systems. This comparative analysis explores the distinctions in criminal liability between weak and strong artificial intelligence, considering their varying levels of autonomy and decision-making capabilities.

The study begins by defining weak and strong AI, with weak AI referring to systems that are narrowly focused and exhibit limited autonomy, while strong AI denotes systems capable of general intelligence and independent decision-making. It then delves into the legal frameworks governing criminal liability and encompassing traditional legal principles and legislation.

Drawing on relevant case law, the analysis examines the challenges of attributing criminal responsibility to weak AI. Due to their limited autonomy and reliance on human input, weak AI systems are typically treated as tools rather than independent agents. Consequently, liability is more likely to be assigned to the human actors responsible for designing, operating, or utilizing the AI system, rather than the AI system itself.

In contrast, strong AI presents unique legal and ethical complexities. With their potential to exhibit cognitive abilities akin to human intelligence, strong AI systems raise questions about whether they should be held accountable for criminal actions. The analysis explores possible approaches to determining liability for strong AI, including the adoption of personhood-like frameworks or the establishment of new legal standards specifically tailored to AI systems.

In conclusion, this comparative analysis underscores the need for nuanced approaches to criminal liability in weak and strong AI systems. While weak AI primarily implicates human actors, strong AI poses intricate challenges requiring the development of innovative legal and ethical frameworks. By addressing these issues, society can navigate the evolving landscape of AI technology, ensuring accountability, fairness, and responsible innovation.

Artificial Intelligence, Criminal Law, Autonomous Driving

Abstrakt:

Rychlý rozvoj umělé inteligence (AI) vyvolal důležité právní a etické otázky týkající se možné trestní odpovědnosti AI systémů. Tato komparativní analýza zkoumá rozdíly v trestní odpovědnosti mezi slabou a silnou umělou inteligencí, přičemž bere v úvahu jejich různou úroveň autonomie a schopností rozhodování. Studie začíná definicí slabé a silné AI, kde slabá AI odkazuje na systémy, které jsou úzce zaměřené a projevují omezenou autonomii, zatímco silná AI označuje systémy schopné obecné inteligence a nezávislého rozhodování. Dále se zabývá právními rámy upravujícími trestní odpovědnost a zahrnujícími tradiční právní principy a legislativu.

Vycházejíc z relevantní judikatury, analýza zkoumá výzvy při přisuzování trestní odpovědnosti slabé AI. Vzhledem k jejich omezené autonomii a závislosti na lidském vstupu jsou slabé AI systémy obvykle považovány za nástroje spíše než za nezávislé agenty. V důsledku toho je pravděpodobnější, že odpovědnost bude přiřazována lidským aktérům odpovědným za návrh, provozování nebo využívání AI systému, spíše než samotnému AI systému.

Naopak, silná AI představuje jedinečné právní a etické komplexity. S možností projevit kognitivní schopnosti podobné lidské inteligenci vzbuzují silné AI systémy otázky ohledně toho, zda by měly nést odpovědnost za trestné činy. Analýza zkoumá možné přístupy k určování odpovědnosti silné AI, včetně přijetí rámců podobných osobnosti nebo stanovení nových právních standardů specificky přizpůsobených AI systémům.

Závěrem tato komparativní analýza zdůrazňuje potřebu nuancovaných přístupů k trestní odpovědnosti ve slabých a silných AI systémech. Zatímco slabá AI především zahrnuje lidské aktéry, silná AI představuje složité výzvy, které vyžadují vypracování inovativních právních a etických rámců. Adresováním těchto otázek může společnost zvládat se rozvíjejícím se prostředím technologie AI a zajišťovat odpovědnost, spravedlnost a zodpovědnou inovaci.

Umělá Inteligence, Trestní Právo, Autonomní Řízení