Conclusion and Extended Summary[1954]

The determination of liability in crimes involving autonomous systems driven by artificial intelligence presents numerous challenges. The issue has been a subject of extensive debate in the legal literature in recent years, with diverse opinions being advanced. This study sought to provide concrete solutions for the determination of the liability of 'the persons behind the machine', particularly focusing on negligent liability, within the framework of criminal law dogmatics. While the majority of existing studies tend to concentrate on specific AI applications, such as self-driving vehicles, this study attempted to offer a broader and more comprehensive framework. Accordingly, it began by examining the reasons why the topic requires a separate analysis. Subsequently, it explored alternative liability models, such as the robot's own liability and product liability. Following this, it examined causation issues in crimes involving AI-driven autonomous systems, focusing briefly on intentional liability and then providing a comprehensive analysis of negligent liability. In this context, the duty of care in negligence is examined in detail, with particular attention given to the concept of permissible risk. A calibration model is proposed, suggesting that the degree of care should be determined based on the level of risk and societal tolerance. Furthermore, the problem of many hands and the principle of reliance are analysed, recognising the involvement of multiple actors in offences caused by such systems. The widely debated *dilemma* scenarios in the literature are also addressed, and an alternative approach is proposed. Finally, recommendations for *de lege ferenda* are presented.

The concept of '*autonomy*' rather than '*artificial intelligence*' has been emphasised in this study. This choice is based on the rationale that, from a criminal law perspective, the primary issue lies in the (technical) autonomy of these systems, the reduced human control over them, and their potential to generate outcomes that are difficult to predict in advance. Indeed, in the future, AI may evolve differently, change, or the current hype may diminish; even different autonomous entities, including those that are not silicon-based and not currently considered as AI by today's standards, may emerge. In such cases, the findings of this study can also be applied to those

[1954] A detailed examination of the debates, along with specific references to the relevant literature, is provided under the corresponding sections.

autonomous beings, provided that a degree of control remains in human agents.

As with many narratives of humanity, the theme explored here is also timeless, focusing not merely on AI as a novel concept but on the broader notion of autonomy of other beings itself. It can be observed in *Automatons* built by *Hephaestus, Golem* from Jewish folklore (16th century), *Frankenstein's monster* in *Mary Shelley*'s novel from 1818, and many others. Yet, for the first time in modern age, humanity is closer than ever to surrendering control to other entities. Consequently, we are no longer confronting mere puppets; instead, we are engaging with *Pinocchio*, a figure who has transcended his strings. Indeed, with reference to *Carlo Collodi*'s celebrated tale of "*Pinocchio*", unlike simple mechanical dolls, *Geppetto* does not have total control over *Pinocchio*. In fact, due to his unpredictable temper, all *Geppetto* can do is try to teach him good manners and discipline, just as humans endeavours with robots. The diminishing degree of human control and the unpredictable nature of AI-driven autonomous systems pose challenges regarding the attribution of harmful consequences caused or influenced by such systems. Therefore, the question becomes: to what extent can *Geppetto* be held liable for the crimes caused by *Pinocchio*?

Among the primary legal challenges arising from the integration of AI-driven autonomous systems into daily life are two fundamental issues, which can be analysed from both *ex ante* and *ex post* perspectives. Leading the *ex ante* challenges is the concept of "autonomy risk" which encompasses unpredictable behaviour and a reduced level of human control over outcomes. Indeed, increasing autonomy and unpredictability of AI-driven systems significantly complicate the analysis of criminal liability for the person behind the machine. These systems possess the ability to make goal-oriented decisions and adapt their behaviour in unfamiliar or dynamic environments without human intervention, relying on advanced "self-learning" and data processing techniques. This complexity (although desirable for the system's success) makes attributing liability more challenging, due to the unpredictability of these systems and the diminishing clarity of human involvement in the causal chain.

Despite the extensive philosophical and metaphysical background of the concept of autonomy, this study adopts the established notion as it is represented in the legal and technical literature. Accordingly, a system can be considered to exhibit (technical) autonomous characteristics if it is capable of performing specific tasks independently of direct human intervention. However, it should always be borne in mind that autonomy is not an absolute state but rather exists on a spectrum. In this regard, it is essential to emphasise that these systems differ from automation processes that produce pre-defined outputs, regardless of their complexity. Since the outputs of automatic systems are largely predictable, they generally do not pose significant challenges in terms of liability. On the other hand, the functioning of AI is not akin to magic. While AI-driven systems rely on complex mathematical formulas, statistical methods, and vast datasets, they stand apart from automated systems due to their ability to generate non-predefined outputs. Enabled by machine learning algorithms, these systems operate based on their own perceptions rather than being limited to user inputs. They can develop their own heuristics, analyse environmental data, "learn" from new inputs and "make decisions" accordingly, which distinguishes them fundamentally from traditional automated systems.

In the context of *ex post* challenges to determining liability, the opacity of AI-driven systems poses a significant issue. While advancements in machine learning and deep learning have greatly enhanced AI capabilities, their increasing complexity often comes at the cost of interpretability. This opacity, stemming from factors such as algorithmic confidentiality, the general public's limited technical expertise, and the intricacy of managing vast datasets and numerous parameters, creates a 'black-box' phenomenon. As a result, establishing a clear causal nexus between input and output, as well as certain behaviour and harmful outcomes becomes highly challenging, thereby complicates the attribution of criminal liability. However, in cases where the operational methods of AI systems can be understood, such as when specific behaviours can be traced to their outputs or external interventions can be identified, a causal relationship can be established.

Nonetheless, the complexity of human-machine interactions and interconnected systems amplifies the risks, such as network failures and vulnerabilities to cyberattacks. Legal challenges further arise in distinguishing between harm caused by design flaws, self-learning capabilities, or manufacturing defects. Given the diverse applications and risks associated with such systems, adopting a universal approach to liability is not feasible. While criminal law may serve as a deterrent in certain instances, non-criminal enforcement mechanisms may be more appropriate in others. Resolving these issues requires a careful balance between societal benefits and potential risks, alongside the consideration of tailored solutions, such as proactively designing AI systems to minimise harmful behaviour.

Autonomous systems driven by AI complicate traditional notions of causality by introducing unpredictable and non-linear elements into the chain of cause and effect. Unlike straightforward automated processes, autonomous systems can "learn", adapt, and generate outputs beyond their programmers' initial aims, which makes it challenging to foresee specific outcomes or pinpoint individual liability. For this reason, instead of directly stating that AI-driven autonomous systems "caused" the harm, the broader term "involved" is used to reflect their role at some point in the causal chain leading to the harm. These systems can be involved in a criminal offence in various ways. By focusing on the role of AI systems in criminal offences and taking into account different perspectives in literature, this study analysed the matter under three main categories: 1- *crimes committed through AI systems*, 2- *crimes committed against AI systems, 3- crimes caused by (with the involvement of) AI systems*. The first category refers to the utilisation of AI-driven systems to support or increase the effectiveness of committing an offence. The second category refers to offences targeting AI systems themselves, exploiting their vulnerabilities or manipulating them in various ways. The third category, which forms the primary focus of this study, encompasses more complex scenarios in which AI-driven systems exhibit autonomous characteristics and human control is limited or even absent.

The study examined more than forty incidents involving AI-driven autonomous systems as illustrative examples under relevant sections. Despite the considerable number of such incidents, particularly those involving semi-autonomous driving, that have attracted media attention in recent years, there have been almost no criminal law cases to date (apart from a few cases in the U.S.) that examine the issue through concepts such as the principle of guilt, individual criminal liability, the scope of the duty of care, permissible risk, and the principle of reliance.

Because of their inherent autonomy and opaque nature, criminal liability in cases involving AI-driven autonomous systems poses significant challenges, leading to what the literature describes as a "liability gap" in criminal law, that existing legal frameworks struggle to address effectively. To address this issue, certain liability models have been proposed in the literature. The first of these is the recognition of legal personhood for robots and holding them liable. Indeed, the question of whether AI-driven autonomous systems should be granted legal personhood has given rise to significant debate. Proponents of this idea, often influenced by anthropomorphic perceptions, argue that advanced AI systems should be recognised as legal persons to address liability gaps, citing examples such as corporate personhood and the recognition of other non-human entities to support their position. Some emphasise the increasing complexity of AI and its capacity for human-like interactions, proposing that such systems should, for pragmatic reasons, be held accountable for damages not merely as tools but as agents capable of bearing responsibility. The opposing viewpoint, on the other hand, highlights that the absence of free will and moral agency (both of which are fundamental aspects of criminal liability) is a limitation inherent in AI. Even the most sophisticated AI systems are incapable of engaging in genuine moral reasoning or comprehending the consequences of their conducts, which precludes their suitability for criminal liability. European legal traditions, which are grounded in individual culpability, are reluctant to extend personhood to non-human entities. They also express concern that attributing liability to AI-driven systems may result in the evasion of liability by persons behind the machine, which would be inconsistent with the core principles of justice.

In my opinion, all arguments for recognising personhood in robots, apart from those based on pragmatic necessities, are inherently contradictory or misrepresent the essence of the concept. Mainly because they fundamentally lack genuine moral reasoning, a will and the capacity to understand their conducts, it is not feasible. Even adopting a pragmatic or functionalist approach to grant personhood to AI-driven systems through a fiction presents significant challenges, particularly in determining which entities should be eligible. One might argue that legal personhood could only be granted to those registered in an official registry. However, the wide variety of AI systems, from simple software to advanced deep neural networks, complicates the issue, as these systems can be easily created, divided, and reassembled. Such systems are unlikely to possess an actual will; however, what is presently observed is an illusion of one. As machines advance and demonstrate increasingly sophisticated capabilities, this illusion becomes more convincing. Nevertheless, it remains fragile; even a minor error can easily disrupt this perceived impression of will. Another fundamental reason why AI-driven autonomous systems cannot bear their own liability is their inability to perform a legally valid act. The matter has been examined in detail in the study. Consequently, although some perspectives in the literature from the Anglo-American legal tradition, argue that robots could fulfil the elements of *actus reus* and even *mens rea*; it is not possible to assert that robots can perform actions in the sense required by criminal law, according to existing theories of action. According to one perspective, the content of concepts can evolve over time, and the concept of action in criminal law could adapt to address the unique challenges posed by robots, considering their rule-based programming as an alternative form of volitional conduct. It can be argued, on the other hand, acknowledging that language is a living phenomenon and that concepts evolve over time, the primary question that must be addressed is whether it is truly necessary to hold robots liable. Criminal law, along with its concepts and principles, was developed specifically for human beings. Therefore, applying these concepts to different entities through reinterpretation could lead to entirely new and complex problems. Even if such fictions are created, they may contradict with real-life practices. Therefore, should such a necessity arise in the future, rather than adapting or extending our current legal constructs to accommodate these circumstances, we would require an entirely new legal framework, or even paradigm.

Focusing on the "liability gap" which is highlighted in the literature, and considering the difficulties in determining criminal liability and attributing it to a specific individual, the study examined how offences caused by AI-driven autonomous systems are addressed through other forms of liability and analysed whether these approaches can be adapted to criminal law. First, a comparison of fault-based liability has been conducted to highlight the differences between civil law and criminal law. Civil and criminal law share certain foundational elements related to fault, but they differ significantly in their purpose and application. Civil law primarily aims to compensate the injured party, permits strict liability, and often adopts a different degree for standards of care, facilitated by the insurability of risks. In contrast, criminal law focuses on punishing personal wrongdoing, requires negligence to be expressly prescribed by law, and prohibits strict liability under the principle of *nulla poena sine culpa*. Moreover, despite differing views in the literature, the concept of negligence differs between the two fields, as they serve distinct purposes.

The existing literature has sought to address offences involving autonomous systems, which push the boundaries of traditional criminal law dogmatics, by analysing similar phenomena to develop potential solutions. In this regard, some scholars draw analogies between AI-driven autonomous systems and concepts like slavery, animal ownership, or employer-employee relationships; arguing that, just as a master or employer might be liable for the actions of a slave or employee, those who control AI should similarly bear responsibility for AI-generated harms. Historical doctrines such as *respondeat superior* and *noxal liability*, which attribute liability to individuals with a supervisory role or beneficial interest, have been analogised to justify imposing vicarious liability on AI developers or owners.

However, this approach falls short in criminal law, as criminal liability requires personal culpability, which cannot be fulfilled solely by occupying a supervisory role. Furthermore, to address the challenges of fault-based liability in offences involving AI-driven autonomous systems, it has been proposed to adapt strict liability in criminal law to fill "liability gaps" and ensure accountability for harm that might otherwise be dismissed as "bad luck". Although this approach may be applicable in other legal traditions, it is largely flawed within the framework of the Continental European legal tradition, where culpability remains a fundamental cornerstone of criminal liability. Thus, the strict or vicarious liability models seen in civil law, conflict with foundational principles of criminal law, and therefore cannot be straightforwardly transposed onto criminal liability for AI-driven systems.

Consequently, after establishing that robots cannot be subject of liability and that civil law liability models are inadaptable into criminal law, the likelihood of many offences involving AI-driven autonomous systems not being penalised becomes increasingly apparent. While such issues may be addressed by civil or administrative law, it is argued that a criminal liability gap has emerged. However, a purely compensatory approach may fall short of meeting society's expectations for justice and may weaken the perceived legitimacy of the legal system. In the absence of punitive or deterrent measures, civil law remedies are inadequate, and even potential compensation fails to function as a real deterrent when absorbed by industries or insurers that can incorporate them into their calculations in advance. Humans are often driven by a retributive sense of justice, and such approaches solely aiming to deter future offences are insufficient. In a future where robots undertake the majority of tasks, it is crucial to consider how the existence of a "retribution gap" rather than merely a "criminal liability gap" will impact society. In other words, the deployment of sanctions in other domains of legal practice to address infringements may result in a retribution gap that can only be addressed through the mechanisms of criminal law. Thus, from the standpoint of legal dogmatics and policy, the question becomes: in the event of a fatal multi-vehicle accident caused by a self-driving taxi, will the families of the deceased truly feel that "justice is served" by a sincere apology from the manufacturing company and compensation in the form of a five-figure sum in US dollars, when no one can be held criminally liable? Therefore, solutions must be developed to address society's retributive needs adequately; otherwise, they will be disregarded altogether.

The study examined product liability as a viable model, which holds particular significance in the context of AI-driven systems, whose increasing autonomy diminishes user control while the characteristics of these systems are predominantly determined during the training and production phase. Consequently, the role of manufacturers becomes even more critical. In civil law, product liability, which predominantly takes the form of strict liability, can be applied to AI-driven systems. However, three main issues arise in the context of product liability for AI-driven systems. First, there (was) the challenge of defining AI as a 'product' within this framework. Second, the interpretation and scope of 'defect' in AI-driven autonomous systems requires careful analysis, since traditional definitions may not encompass the unique, evolving characteristics of such systems, in particular for adaptive, "self-learning systems" which have the capacity to evolve even after reaching the end user. And third, the burden of proof poses significant challenges, particularly given the inherent opacity of many AI systems.

Criminal product liability, unlike its civil counterpart which primarily seeks compensation for harm, requires proof of individual fault and focuses on punitive and deterrent objectives. Therefore, it imposes a stricter evidentiary burden in establishing causation and individual wrongdoing. The development of criminal product liability, assessed within the framework of existing criminal law in the absence of a distinct positive legal regulation, has been significantly shaped by the German Federal Court of Justice (BGH). The responsibility of manufacturers within this framework can be summarised as ensuring the marketing of adequately tested and safe products; informing users about proper use, existing and potential risks; actively monitoring the product and taking necessary measures, including recalling the product if suspicions arise regarding its harmful consequences arising from the guarantor position. The determination of criminal product liability involves, first, examining whether the manufacturer has engaged in any conduct subject to assessment under criminal law, through the product. Following this, the behaviour of the individual employee or board member is examined within the framework of their duty of care. Furthermore, it should be noted that the BGH has introduced a different approach in light of the unfeasibility of definitive scientific proof of the outcome.

Intentional crimes will constitute exceptional cases in the context of AI-driven autonomous systems. Such crimes, when committed by employing these systems, are largely treated as if the AI was merely a tool or instrument, akin to a dog or a piece of equipment used to cause harm. Although the exact outcomes of such actions may not always be foreseeable *ex ante*, this is comparable to a situation where a person who uses poison to kill another does not need to know the precise effects of the poison. In cases where the outcomes of AI-driven systems are generally foreseeable, intentional liability will arise.

In criminal law literature, a significant number of scholars have argued that the indirect perpetration model can be applicable in cases where AI-driven autonomous systems are utilised to commit criminal offences. However, I hold the opposite view, arguing that it is inapplicable in such intentional offences; mainly because theoretically, the indirect perpetrator utilises not another person's physical body but their actions as a tool, through exercising control over their will. In this regard, it is not possible to invoke indirect perpetration in cases where AI-driven autonomous systems are utilised to commit crimes, because: (1) they lack will; (2) their conduct cannot be considered an act in the sense of criminal law, and (3) they are not human to be considered as "another". Even if the requirement for the innocent agent to be human were ignored, and it was accepted that AI-driven autonomous systems could perform acts in the sense of criminal law; they would still need to possess a certain level of will for this debate to hold any meaningful relevance.

The majority of offences involving AI-driven autonomous systems are likely to pertain to negligent crimes. Despite the unpredictable outputs of these systems, numerous measures can be implemented during the training phase and after deployment to ensure mitigating their risks. The major challenge in negligent liability for AI-driven autonomous systems is that, although manufacturers and developers retain some control during design and updates, they cannot fully predict or prevent every harmful outcome once the system is deployed. Additionally, because users also influence the system's operation, the distribution of responsibilities becomes blurred, which makes it difficult to establish foreseeability and pinpoint the precise causes of harm.

In criminal law, establishing the source of the duty of care and defining its scope and boundaries is essential in the cases of negligent liability. The duty of care derives from a multifaceted framework encompassing statutory legal provisions, behavioural standards, codes of conduct, professional guidelines, administrative and operational instructions, usage protocols, and unwritten norms. Additionally, where necessary, it requires adherence to the *state of the art*. Furthermore, when engaging in potentially risky activities, the general principle of refraining from harm is also applicable. Therefore, merely ticking boxes by complying with written norms may be insufficient; a comprehensive approach to risk mitigation is required. A significant issue concerning the state of science and technology in AI-driven systems is that this field, due to its substantial investment requirements and inherent risks, is led by a small number of large corporations. Typically, the entities advancing the state of the science and technology are the same companies developing these products. Consequently, these companies must not only bring such products to market but also continue to develop methods to minimise their associated risks. They must not abandon research and development efforts to evade liability. Legal systems should adopt measures to ensure the continuation of such efforts. Ultimately, whether the duty of care has been fulfilled will be determined by the courts based on the specific circumstances of each case.

Whether negligence should be evaluated by a general and objective or individualised standard of care has been an important point of discussion. The two-stage analysis of negligence, the individualisation theory and other perspectives offer distinct frameworks for the evaluation. The study examined the issue in detail, demonstrating that different theoretical frameworks often take divergent paths yet ultimately arrive at similar practical outcomes, although opposing views do exist. A central debate in determining a breach of duty of care is whether the perpetrator's specialised knowledge and skills, or their general incompetence, should be considered; with the prevailing view asserting that those with greater expertise should be held to higher standards of care. Nonetheless, imposing higher standards may inadvertently discourage companies from acquiring advanced skills or knowledge by subjecting them to greater obligations. Additionally, it could deter them from conducting comprehensive risk analyses or investigating emerging technological risks. To address this issue, it would be prudent for the legislature to explicitly impose such obligations, thereby fostering a proactive approach to the identification and management of potential risks.

The prevailing opinion holds that special abilities and knowledge should also be taken into account. For instance, if a programmer employed by a company discovers that the company's AI system, *e.g.* a large language model (LLM) processes confidential state secrets and discloses them in response to ordinary user queries, it would be unreasonable to expect a programmer to remain silent and merely continue performing their regular duties. Similarly, if a method to reasonably mitigate the risks associated with a self-driving vehicle is identified through research conducted by a specific company, but this method has not yet become an industry standard and is not implemented by other companies, the company in question is nonetheless obligated to adopt the method to reduce the risks. Failure to do so could result in criminal liability.

On the other hand, below-average abilities cannot exempt an individual from liability. While criminal law generally takes into account the offender's personal attributes and abilities under the concept of culpability, individuals who lack the personal capacity to meet the objective standard may still incur liability if they willingly undertake a task for which they are unqualified. Thus, negligent undertaking occurs when an individual, despite lacking the requisite competence, engages in a risky or complex activity and thereby fails to maintain the necessary level of care. The practical implication of this concept is that, particularly in the context of high-risk systems, only a limited number of highly advanced companies may be able to operate. While this might appear to be a positive outcome, it carries significant risks, particularly given the strategic nature of certain sectors and the potential for these companies to impose their own biases. Another aspect concerns the use of self-driving vehicles, which, while facilitating mobility, particularly for individuals with physical limitations, may occasionally require human intervention. If such vehicles are used by individuals incapable of taking control when necessary, this could constitute negligent undertaking. To mitigate this risk, it may be prudent to require users to complete a training course before being allowed to operate these vehicles.

In the context of negligent liability, the scope and boundaries of the duty of care are of critical importance. The duty of care encompasses considerations such as foreseeability, adherence to established standards, risk mitigation, proactive prevention, reasonable behaviour, awareness, and the avoidance of omissions where action is required. For a breach of the duty of care to be established, the harmful outcome must have been both foreseeable and avoidable. However, when it comes to AI-driven autonomous systems, their "self-learning" capabilities and adaptability make foreseeability, and more broadly, the ability to recognise potential outcomes, particularly challenging.

The boundaries of foreseeability have been extensively discussed throughout the study. In my view, it is incorrect to claim that liability cannot arise merely because the outputs of such systems are deemed unforeseeable. Indeed, these systems inherently carry certain risks, and the unforeseeability of the typical risks posed by AI-driven autonomous systems is itself recognisable. For instance, in the case of a tiger released from a zoo, the risks it may pose are broadly recognisable: it might attack a few passers-by. It is, however, unlikely to simultaneously bite 100 individuals, cause a plague, or compromise personal data. In other words, typical risks are generally recognisable, and the inability to control such systems at every stage, as if they were puppets, does not negate this fact. The introduction of these systems, along with their inherent risks, serves as the foundational anchor point for analysing liability. Therefore, the point of inquiry for assessing liability should centre on the moment a task is delegated to an AI-driven system. This does not imply that liability will arise in every instance. Indeed no one can be held liable for matters beyond their control. However, the key point being emphasised here is that, within the framework of criminal law, the focus should be on the act related to the use of such systems at the time it is performed. Subsequently, other factors will be assessed to determine liability. In this regard, issues such as identifying whether the risk has been enhanced or mitigated are of critical importance. A manufacturer's defence based on the claim that potential harmful outcomes were unforeseeable should instead shift towards an obligation to identify, and where possible, reduce the risks. In other words, rather than focusing solely on the foreseeability of harmful outcomes, potential dangers must also be researched and recognised.

Autonomous systems driven by AI can produce unexpected, almost unforeseeable outcomes, some of which may be classified as 'black swan' events. Nevertheless, it is crucial to draw lessons from such incidents and adjust the standard of care to reflect these experiences in subsequent assessments. Therefore, it would not be incorrect to assert that the duty of care possesses a dynamic and evolving nature. For example, prior to 2015, it may not have been reasonable to expect developers of robot vacuum cleaner software to anticipate and design their systems to recognise people sleeping on the floor and prevent incidents such as pulling human hair. However, this has now become part of the duty of care. That said, caution must be exercised to avoid *hindsight bias* in specific case assessments. Moreover, when determining the scope of an individual's duty of care, new possibilities and advancements in technology must also be considered alongside past incidents. For instance, in the *Aschaffenburg case*, it could be argued that in 2012, the absence of a technical system capable of taking over driving and safely manoeuvring a vehicle in the case where the driver lost consciousness, was understandable. However, given the advancements in modern driving assistance systems and semi-autonomous features, such functionality is now expected to meet the standard of care.

The outcome is objectively foreseeable if a reasonably prudent person from the perpetrator's environment under the given circumstances based on general life experience would have expected the occurrence *ex ante*. On the other hand, objective foreseeability is rejected if the occurrence of the outcome is so far from everyday experience, such as in cases involving an unusual and improbable sequence of events, that it could not reasonably have been anticipated by no one, including the perpetrator. Foreseeability, particularly in the context of emerging AI technologies, is inherently abstract, and general life experience is of limited relevance. While absolute prediction of every potential outcome is unfeasible, the law expects from the persons behind the machine to recognise typical or broadly predictable risks, distinguishing them from atypical events that lie entirely outside ordinary experience. Yet, typical risks do not always indicate the existence of objective foreseeability, nor do atypical risks necessarily mean that the outcome is absolutely unforeseeable. Nonetheless, requiring absolute foresight would effectively impose a standard of strict liability. In this regard, identifying what constitutes typical risks is crucial. For example, a self-driving vehicle causing an accident due to an improper lane change is a typical risk, whereas its software hacking an information system is atypical. However, distinguishing between typical and atypical risks will require significant time and experience.

In determining whether the duty of care has been fulfilled, reliance on a hypothetical *careful person* standard is also not feasible. This approach carries the risk of excessive generalisation, and moreover, such a standard has not yet been firmly established in AI-driven systems. Indeed, in the context of these technologies, what constitutes diligent behaviour and the applicable standards of conduct remain unsettled. As mentioned, the duty of care arises from a multifaceted framework that includes written legal provisions, norms of conduct, professional guidelines, administrative, operational, and usage instructions, as well as unwritten norms. In this regard, existing codes of conduct, relevant legal and industry standards (such as those regulating autonomous driving) or other standards such as ISO and DIN can also be taken into account. However, fulfilling these serves only as an indicator of compliance with the duty of care. Furthermore, the duty of care is dynamic in nature and may be influenced by factors such as an increase in risk or failures within the system. Moreover, the system must be designed to be robust, ensuring that it is protected against hacking and other forms of interference by third parties. When determining liability in a specific case, it is essential to consider the protective purpose of the norm and whether the harmful outcome resulted from the increased risk. And in any case, the general principle of the duty to refrain from harm applies.

It is a fundamental concept in risk perception that no human behaviour is entirely free of risks nor is any (technical) system without flaws. Enhanced diligence and meticulous attention can serve to mitigate risks, diminishing both the probability and the magnitude of potential harm. However, the complete elimination of all risks is unattainable, even in the most carefully conceived and executed behaviour. Building on this premise, to balance societal needs and risk management, the permissible risk doctrine emerged in the 19th century and was conceptually developed in the first quarter of the 20th century. Therefore, certain risky actions, despite their risky nature are considered permissible if appropriate safety measures and due care are observed. These actions, though inherently dangerous, do not lead to criminal liability as long as the necessary precautions are taken. There are debates in the literature regarding the legal nature of permissible risk. In line with the prevailing view, this study focused on evaluating the limiting effect of permissible risk on the duty of care within this context.

Manufacturers are obligated to research and implement new findings that can identify and mitigate previously unknown risks, thus new methods to identify and mitigate them; reduce their impact or decrease their frequency can be developed. Therefore, in innovative areas such as AI-driven autonomous systems, instead of relying on generally accepted rules of technology (which are not fully established), the continuously evolving and dynamic state of science and technology should be applied to mitigate risks as much as possible. Despite all necessary care being taken, including rigorous testing protocols, continuous monitoring, real-time data analysis, and regular software updates, if users have been warned about both existing and potential hidden dangers, and if no alternative measures to mitigate harmful effects were feasible, the elimination of the remaining risks cannot reasonably be expected. What remains are residual risks, which are considered permissible. Accordingly, if a harmful outcome could have been averted by adhering to the relevant safety regulations, or the general duty of care, the perpetrator cannot invoke the inability to prevent the accident as a valid defence. Furthermore, even within the scope of permissible risk, strict liability under civil law remains applicable.

To illustrate, evidence demonstrates that relying solely on camera-based computer vision in self-driving technology is inadequate. Designing autonomous driving systems with such limitations, driven by economic or aesthetic considerations, cannot be regarded as fulfilling the duty of care to mitigate risks associated with a particular activity. Furthermore, such activity cannot be classified as falling within the scope of permissible risk.

Even in the absence of an established industry standard on this matter, such dangers arising from the product must be prevented where it can be reasonably achieved. Therefore, in a specific incident, if it can be proven that the use of additional sensors, such as LIDAR, would have prevented an accident, the manufacturer may be held liable. The criteria here is whether the failure to employ available technology increased the level of risk in a legally disapproved manner. The defence that autonomous driving is generally safer than human drivers is insufficient. Although the failure to implement these methods may not be evident in individual cases, it would statistically increase the number, type, and severity of accidents. Thus, avoiding the use of readily available technologies capable of preventing accidents, solely for aesthetic or economic reasons, gives rise to liability for negligence.

Permissible risk doctrine does not provide a *carte blanche* and that only certain risks can be deemed permissible under strict conditions. The question arises whether atypical risks can also be considered permissible. Undoubtedly, determining whether a risk is typical requires experience-based data, which is not yet available for AI-driven autonomous systems. The resolution of this issue is not adequately guided by the concepts of protective purpose or *ratio legis* of the norm, or legally relevant risk, either. For instance, one might consider a hypothetical scenario where a self-driving bus fails to correctly classify a child disembarking from the vehicle, leading to the vehicle's door trapping the child's hand, causing injury. In such a case, it is difficult to argue that this injury should fall within the scope of permissible risk merely because self-driving vehicles are expected to significantly reduce traffic accidents. Consequently, it is not readily apparent that society should tolerate incidents of this kind within the broader framework of permissible risks. In the context of negligent liability, the key issue to be assessed here is whether adequate and necessary testing and safety measures were implemented to prevent such a failure of the door. Similar discussions can be applied in cases where the software of a self-driving vehicle hacks into an information system. Therefore, instead of distinguishing whether a risk is permissible based on typical and atypical risks, the distinction should be made based on the recognisability of the causal chain. Nonetheless, in areas where risks are not fully recognised, such as AI, it remains important to identify the atypical risks.

In this study, discussions on permissible risk and social adequacy in the context of sports competitions are included to better address the unforeseen outcomes, and the distinction between typical and atypical risks. In light of the explanations and past scholarly debates on legal background of sports, it can be stated that recognising atypical risks under permissible risk doctrine or considering them socially adequate is difficult. Indeed, permissible risk in sports encompasses the typical risks of the activity as long as the rules are adhered to (or in cases of minor breaches). However, in situations where the degree of harm significantly increases, the explicit consent of the affected party may be additionally required. Intentional or harmful behaviour outside the flow of the game is strictly prohibited. In this regard, it can be argued that for certain atypical risks posed by AI-driven autonomous systems, the explicit consent of the affected individuals could be sought. Such consent would be legally effective only if it fully satisfies the detailed conditions for valid consent under the law. However, this approach would only be applicable in extremely limited circumstances, as many AI-driven autonomous systems cause harm to uninvolved third parties without the possibility of obtaining prior consent. Moreover, the extent of such harm may be of a nature that cannot be consented to. In such cases, while the invocation of presumed consent might be considered, in my view, this would also be inapplicable. For instance, a person deciding to use a robotic vacuum cleaner would likely not consent to being injured by having their hair pulled if asked beforehand.

Therefore, as determining typical and atypical risks in emerging technologies requires time and experience, the scope of areas left unpunished -particularly those involving serious consequences such as harm to life and limb- should be kept extremely limited. In the assessment of a risk as socially tolerable, in addition to its societal gains; objective and verifiable criteria, such as the severity and extent of the damage, its probability and proximity of occurrence, the rank and value of the affected legal interests, available prevention and control options, and whether the damage is irreversible, should play a central role. The need to safeguard societal safety while avoiding excessive restrictions that could hinder innovation should not be addressed through a balancing of interests akin to that employed in cases of necessity. Such an approach would introduce a utilitarian framework into the permissible risk doctrine, which is particularly problematic in scenarios where human life is at stake.

This study adopted a risk-based approach regarding the permissibility of risks. Accordingly, the duty of care to be applied should be calibrated by balancing the societal significance and necessity of the activity in question against the level of risk. This calibration is grounded in two prior works from German legal literature, which classify risks into specific categories.

Based on these classifications, the risks posed by AI-driven autonomous systems, their societal benefits, and the extent to which their risks can be mitigated through due care are schematically analysed to establish a framework for calibration. For instance, an AI-driven system or activity that serves only a limited number of individuals and provides no meaningful societal contribution is classified as socially useless. Such systems are permitted only in cases of low to moderate risk, provided that high levels of care are exercised. If the activity involves high risks, it is not permitted unless those risks are significantly mitigated. Conversely, an AI-driven system deemed socially useful is subject to varying levels of care based on the degree of risk it poses: low-risk systems require a lower duty of care, whereas high-risk systems necessitate a increased duty of care. In essence, the aim is to establish a reasonable and practicable framework by determining the necessary duty of care in proportion to the societal benefits of the activity and the associated risk levels.

For an activity to be considered within the scope of permissible risk, the inherent risks of that particular activity should be tolerated by society. This societal tolerance is typically evaluated by balancing the activity's social utility and benefits against the level of risks involved. In this regard, a significant issue arises when one party benefits from a particular activity or technology, while another, whose interests are infringed upon through exposure to it, suffers harm. Therefore, the permissiveness of the risks must be grounded in clear and well-defined basis, whether it stems from societal consensus, public interest, or another appropriate framework. There must be a transparent and inclusive discussion about the advantages of these systems, identifying both the beneficiaries and those who bear their risks. If the system endangers entirely uninvolved parties, the permissible scope of risk should be minimal. Conversely, if users or others knowingly and voluntarily accept the associated risks, the threshold for permissible risk may be correspondingly higher. In this context, it can be stated that social adequacy (*soziale Adäquanz*) reflects societal acceptance of certain risky behaviours over time on various grounds and serves as an interpretative tool rather than a determinant of permissible risks.

The extent to which society is willing to accept and tolerate the risky activity is of paramount importance. It is possible to propose certain points on this matter. First, society's perception of risk is inherently subjective, and there is a notable lack of objective empirical data, particularly longitudinal studies, on the real-world testing of AI-driven autonomous systems, including their actual dangers and benefits. Secondly, if the risks of a particular activity are to be tolerated by highlighting factors such as its contributions to the environment and the economy, the other side of the coin must also be considered; namely, its overall negative impacts. Therefore, it is also crucial to consider the irreversible delegation of control from society to autonomous systems. As can be observed, in the desire to use of autonomous taxis, the process begins with the delegation of specific tasks but is likely to evolve in the near future into the delegation of almost all activities in smart cities, leading to a significant diminution of human control. Third, emerging technologies, not only facilitate tasks previously undertaken by individuals but also gradually become new societal norms, thereby increasing the scope of personal and social responsibilities over time. In cost-benefit analyses, this phenomenon, which unfolds over time, is often overlooked. Fourth, it would be naive to suggest that this process unfolds within a framework of conscious and deliberate societal debates. In practice, fundamental rights and freedoms are often irreversibly altered through the interplay of rapid societal dynamics, advancing technology, and those who control (and benefit from) it. A pertinent example is the swift abandonment of privacy in the face of rapidly progressing technological developments. Rather than society willingly accepting its risks and drawbacks, the use of smartphones for instance, has become a necessity, imposing itself as an indispensable part of daily life. Finally, in evaluating the acceptability of risks, balancing society's various interests is crucial; however, it must be borne in mind that different segments of society may have divergent interests, and the paramount consideration should always be the general benefit of public.

The direct societal gains and potential dangers of these systems play a significant role in the societal acceptance of the relevant activity with its inherent risks. For example, one of the most prominent examples of robotics, self-driving vehicles, claim to offer numerous advantages, including enhanced safety through the reduction of human error, increased mobility for individuals unable to drive, and improved traffic flow, which helps reduce emissions and congestion. Additionally, AI-driven autonomous systems are successful at undertaking dangerous, repetitive, or specific tasks, such as deep-space exploration or detailed medical image analysis. They deliver greater efficiency and in some cases reliability than human operators, thereby mitigating physical risks and time constraints. Furthermore, by processing vast quantities of digital information, they enable intellectual collaboration across different fields and support human judgment where critical decision-making is required.

On the other hand, AI-driven autonomous systems pose several significant risks, including vulnerabilities arising from network interconnectivity, privacy intrusions due to extensive data collection, and the reduction of human oversight and control. The opacity of complex AI models can obscure accountability and perpetuate harmful biases, particularly when these systems are trained on flawed or discriminatory historical data. Moreover, excessive reliance on AI tools may degrade the quality of outputs over time, especially if newer models are trained on the often-average results of earlier systems that rely on synthetic data. These risks can become even more amplified when they transition from isolated threats, such as hacking a single device, to large-scale issues, such as the coordinated manipulation of networked vehicles, ultimately jeopardizing not only individual interests but also entire social infrastructures. Indeed, while the networking of systems already poses numerous risks, the autonomous features of interconnected systems aggravate these risks. For instance, whereas a malfunction might typically occur in a single unit, erroneous "learning" processes or flawed software updates can result in mass malfunctions. In scenarios where such systems are deeply integrated into societal functions, these failures can lead to significant and uncontrollable disruptions. Furthermore, the displacement of human labour and the erosion of essential human judgment raise profound ethical dilemmas, including the potential dehumanisation of decision-making processes and the undermining of core societal values. Consequently, societies must carefully balance these potential harms against the benefits, adopting well-calibrated oversight and regulatory frameworks that address both immediate risks and the broader systemic transformations brought about by AI.

The dominant approach in the literature focuses on evaluating whether the AI-driven systems offer greater safety compared to human execution of a particular task. In the study, however, when evaluating the permissibility of risks, it is emphasised that risk is not a quantitative factor that inherently increases or decreases when a concrete task (such as driving) is delegated to an AI-driven system; rather, existing risks are transformed and substituted with new ones. For instance, while self-driving vehicles may generally reduce the overall number of accidents, they have also been involved in numerous fatal and injurious crashes resulting from simple errors that humans would likely never make. Admittedly, it can be argued that such incidents will decrease as technology advances. However, the point being emphasised here is not limited to self-driving vehicles but extends to AI-driven autonomous systems in general, highlighting that they possess both advantages and disadvantages and even within a specific activity, they may reduce certain dangers while simultaneously increasing others. Therefore, when discussing the societal acceptance and tolerance of permissible risk, it becomes evident that a holistic perspective is required; one that considers all the factors altered by the replacement of existing methods with the new technology. From this comprehensive standpoint, the critical issue lies in the delegation or transfer of a given task to AI-driven autonomous systems.

In other words, the moment of delegating a task and dangers of a particular activity marks the starting point of liability analysis. Following this, it can be assessed whether delegating a task to AI-driven autonomous systems instead of relying on conventional methods introduces new risks, enhance existing ones, or enable the task to be performed with reduced risks. Accordingly, it is inaccurate to assert that such risks are entirely uncontrollable or unforeseeable. Emphasising once more, the moment of delegating control over the relevant task to these systems should serve as a starting point for liability analysis. Although many of these products are generally regarded as safe(r), during their initial stages of adoption, they often bring about a range of unrecognised risks. The prevailing view on this matter seeks to determine whether the harmful outcome would have occurred even if the alternative conventional method had been chosen, and whether such an outcome was unavoidable. In the absence of such certainty, this view advocates the application of the principle of *in dubio pro reo*. Another view (*Risikoerhöhungstheorie*), which I also endorsed in this study, however, examines the situation based on whether the risk was enhanced and attributes liability accordingly. Indeed, particularly with AI-driven autonomous systems, the challenges of *ex post* analysis make achieving certainty unfeasible. In the context of new and particularly high-risk activities, delegating responsibility and liability of a task to such systems demands caution, especially when it risks violating significant legal interests of uninvolved parties. If the use of these systems results in a higher likelihood or greater severity of harm to legal interests, or if the significance of the legal interest at stake increases, the negligent liability may come into question. In this regard, excluding liability where it cannot be definitively proven that the outcome would have still occurred using conventional methods could create a significant liability gap concerning AI-driven systems, whose outputs are often opaque.

In this regard, contrary to the widespread opinion, this study suggested adopting a cautious approach to immediately classifying certain risky activities as falling within the scope of permissible risk and viewing individuals as entirely passive in such scenarios. Indeed, such individuals create a risk by activating the vehicle for example when commuting to work, and delegate a task to the AI-driven autonomous system that is inherently risky. For instance, a person who opts for autonomous driving instead of driving their vehicle on a particularly snowy day might actually increase the existing risk. By avoiding the risk entirely, they may effectively evade liability. Indeed, it has been emphasised in the literature that clever offenders may exploit the permissible risk doctrine. Legal systems should approach such situations cautiously and refrain from generalising that "autonomous driving will generally result in fewer fatalities". Unless the individuals are entirely passive throughout the whole process, the moment of activation or delegation of a task should form the basis for liability analysis. This issue is likely to become even more significant in the future as more tasks are delegated to AI-driven autonomous systems. It should be emphasised that the matter is not merely about identifying an individual to hold liable (since criminal law does not seek someone to *scapegoat*); but rather about determining liability arising from delegating certain tasks to robots despite their inherent risks which are recognisable. Whether such delegation falls within the scope of permissible risk must separately be evaluated.

Another emerging issue, which is likely to become more prevalent in the future, is whether the non-use or deactivation of these systems constitutes a breach of the duty of care. As AI-driven systems become safer and more widespread, failing to utilise them may result in liability for negligence, particularly when these systems clearly reduce risks more effectively than traditional methods. Although such developments have not occurred yet, in the smart cities of the future currently being designed, many human activities have the potential to become atypical. For instance, in a city surrounded by networked, interconnected transportation vehicles, a driver wishing to operate their own car may be considered atypical and pose a risk to safety. In such a situation, it might even be argued that this activity constitutes a luxury and may no longer be permitted. This scenario could apply to many AI-driven autonomous systems. From a legal policy perspective, if such a transformation is inevitable, it must occur in a manner that does not conflict with humanity's evolutionary legacy or intrinsic nature.

The scope of permissible risk and the boundaries of duty of care may not always be clearly recognisable *ex ante*. This uncertainty, combined with the potential risk of future liability, may serve as a significant deterrent for individuals and organisations operating in this field, reminding of the image of the *Sword of Damocles*. To clarify the scope of the duty of care, it would be prudent to define it through specific standards and norms of conduct. Indeed, certain legal rules already incorporate references to standard practices within the industry or the prevailing *state of science and technology*. While the *state of science* is often difficult to achieve in practice, the *state of technology* tends to offer stronger concrete measures for risk mitigation. Relying on this criterion, rather than industrial standard practices, is more effective for risk mitigation, as standard practices may be outdated, or higher standards may be entirely avoided by companies for economic reasons. Establishing such concrete benchmarks would also reduce the need for frequent revisions to legal rules, particularly with respect to rapidly evolving technologies, thereby ensuring their continued applicability. Furthermore, it is neither feasible nor practical for official authorities to regulate applicable rules through a detailed and exhaustive list subject to periodic updates. Such an approach would often fail to adequately address the current state of technology and would risk becoming outdated.

However, it must be emphasised that mere compliance with such written standards does not necessarily equate to the proper fulfilment of the duty of care. Such standards often serve merely as indicators. Engaging in a superficial "box-ticking" exercise does not absolve individuals undertaking such risky activities from liability. In any case, the overarching principle of the duty to refrain from causing harm remains applicable. Indeed, the concept of permissible risk does not grant the actor a *carte blanche*. Even when acting within the generally permissible limits, this does not absolve the actor from the obligation to take additional precautions in specific situations beyond what general standards of care require. If the realisation of a risk is foreseeable in a given situation, the actor has a duty to prevent it, provided they are still in a position to avert the harmful outcome. Legally defined standards of duty of care serve as a baseline but are not absolute; they may be exceeded depending on the specific circumstances and the potential risks involved. Fulfilling the duty of care may necessitate a broad spectrum of actions. Consequently, behavioural rules are supplemented, or even overridden, by the overarching principle of achieving the best possible avoidance of harm to legal interests.

The study also examined the EU *AI Act* (AI Regulation) and the *AI Liability Directive*. It concluded that the AI Regulation does not adopt a genuine risk-based approach akin to the one adopted in the study. It does not provide any determinations regarding criminal liability. However, it imposes certain requirements and obligations on the relevant person behind the machine based according to the AI system's risk classification ("unacceptable", "high", "limited" and "minimal" risk). These requirements can serve as indicators under national law for assessing whether the individual has fulfilled their duty of care. Not all obligations or requirements are aimed at risk mitigation, but those targeting risk reduction; such as ensuring human oversight and providing instructions for use, are particularly relevant in the context of criminal liability. Conversely, requirements such as maintaining technical documentation are not directly related to liability for harm caused. Compliance with the obligations and requirements stipulated in the AI Regulation is necessary to avoid criminal liability, particularly for negligent acts, but it is not sufficient on its own. Moreover, adherence to them does not eliminate the requirement to comply with national legal rules.

Autonomous systems driven by AI can lead to harmful outcomes for various reasons. Determining the precise source of such harm, whether it originates from flawed training datasets, programming errors, software or hardware-based malfunctions, or a combination of these factors, presents significant challenges. This difficulty is amplified by the inherent opacity of these systems. Furthermore, the development of AI systems typically involves collaboration among numerous individuals, with layers of code frequently constructed upon pre-existing frameworks. Moreover, the hardware and software of complex systems are often produced and integrated in different organisations. This sophisticated and fragmented process further complicates the attribution of liability for any harm caused.

The challenge of attributing responsibility within complex organisational structures, where numerous individuals contribute in varying capacities to an outcome, is often referred to as "the problem of many hands". This concept is also applicable in the context of the development and use of AI-driven systems, and alongside the principle of reliance *(Vertrauensgrundsatz)*, holds particular significance in criminal product liability. The problem of many hands in the context of AI-driven autonomous systems arises from the extensive involvement of multiple contributors in the design, development, deployment, and oversight of these complex technologies. The allocation of responsibility across numerous individuals and organisational layers creates significant challenges in identifying the specific person or group who should bear liability. The principle of reliance serves to address these challenges by permitting individuals to presume that others will fulfil their respective duties of care, provided there are no evident indications to the contrary. Indeed, its application is not without limits. When it becomes evident that (through concrete indications) it is unreasonable to expect proper or lawful behaviour from others, or where a hierarchical or supervisory duty exists to prevent foreseeable harm, one cannot rely on other parties' conduct. In such circumstances, actors are obligated to act to mitigate risks. Thus, while the principle of reliance facilitates collaborative innovation, it does not absolve individuals in the face of indications of dangers arising during the production, deployment, or operation of AI systems.

With the increasing delegation of tasks to AI-driven autonomous systems, the extension of the principle of reliance to these technologies has become a subject of debate. What is actually meant by this is: (1) whether humans can rely on autonomous and fully automated systems to function correctly, and (2) whether these systems should account for potential human error. While individuals may operate under the expectation that a system which has consistently functioned correctly in the past will continue to do so, evaluating human reliance in machines under the principle of reliance poses significant difficulties. Beyond its theoretical challenges, particularly in today's transitional phase, individuals are expressly burdened with a duty of care that includes the obligation to monitor and verify the proper functioning of these systems. Furthermore, applying the principle of reliance to interactions between humans and machines, or even among machines themselves, is presently impracticable. This is because the conducts of autonomous systems cannot yet be fully predicted or anticipated, which limits the feasibility of extending reliance in this context.

The second question aims to explore to what extent the persons behind the machine, particularly manufacturers, should anticipate and design AI-driven autonomous systems to take potential human errors, misuse and atypical behaviour into consideration; how much of the atypical behaviour could be legally expected, and to what degree is it the manufacturer's responsibility to prevent harmful outcomes? The answer to this question can be framed around the idea that manufacturers should design their products with consideration for users' typical errors. In this context, another relevant issue is the misuse of an AI-driven system by users. This issue can generally be addressed within the framework of the prohibition of regression (*Regressverbot*).

Regarding the extension of the principle of reliance to AI-driven autonomous systems, it can be argued that it is a concept developed to enable individuals to sustain their social lives in harmony. It allows people to avoid the constant burden of meticulously monitoring the behaviour of others and adjusting their own actions accordingly. In contrast to humans, for instance, self-driving vehicles continuously perform risk assessments as part of their operation through their sensors and advanced computing systems, enabling them to manoeuvre in real time. Therefore, it is unnecessary to expect such systems to rely on humans or other natural occurrences in the same manner as humans. While a human driver cannot simultaneously monitor numerous parameters (and therefore, the principle of reliance becomes necessary), a self-driving vehicle can operate with one "eye" on the pedestrian's immediate movements and its other "eyes" (sensors) scanning all other elements of the road environment. Therefore, instead of applying the principle of reliance in its existing form, its content could be adapted to encompass these systems, in accordance with specific application as well.

The study further examined the longstanding ethical *dilemmas* and their legal implications, considering the expectation in the literature that such dilemmas will become increasingly prevalent with the widespread adoption of self-driving vehicles. Indeed, the common belief that self-driving vehicles will inevitably face ethical (and legal) dilemmas requiring them to make critical choices has recently been a subject of significant debate recently. Accordingly, when an accident becomes unavoidable, self-driving vehicles, owing to their processing power, can rapidly evaluate all possible courses of action and select an option. The question of which option the system should choose when confronted with a decision between two negative outcomes, such as those involving the sacrificing of lives, and whether to prioritise quantitative or qualitative values, has been extensively debated in the literature. In this context, the study provided a detailed discussion on whether legal constructs such as *necessity as a justification, necessity as exculpation, the conflict of obligations,* and *supra-legal excusable necessity* offer solutions, especially in life versus life scenarios. In summary, it is concluded that necessity as justification is inapplicable within the framework of the German legal tradition, as life is considered an immutable value, and the criterion of one legal interest substantially outweighing the other cannot be satisfied. Similarly, necessity as exculpation fails to provide a resolution, since the individual(s) responsible for pre-programming the relevant software does not act to save themselves or someone close to them from danger. Although the legal literature includes various debates, particularly regarding symmetrical and asymmetrical danger groups in cases involving individuals who are certain to die, other legal constructs under German law similarly fail to provide a definitive solution in such dilemmas. One of the factors that complicates finding a solution in this context is the fact that complete non-intervention is technically impossible for self-driving vehicles. This is because a collision avoidance system must be programmed, designed and installed to minimise risks, and the absence of such a system would constitute a violation of the duty of care.

Despite the widespread debates on these dilemmas, the study offered an alternative approach by addressing aspects that are overlooked in the existing literature. Indeed, it can be argued that existing perspectives, being overly reliant on traditional moral dilemma debates, disregards a critical point: these dilemma scenarios are thought experiments, and in real-life situations, such absolute certainty (the absolute death of A or B) is seldom possible. In other words, in a real-world "dilemma", not all probabilities of death will be encountered equally. For instance, if a self-driving vehicle calculates a 40% probability of killing one pedestrian and a 98% probability of killing another as a result of its manoeuvre, would this still be considered a dilemma in the traditional sense discussed in the literature? In such a case, would the principle that human life should never be subject to comparison or weighing remain applicable? Moreover, if the probabilities were 2% versus 98%, would it still be argued that these outcomes are morally or legally equal? The optimal course of action in programming self-driving vehicles is to establish a system which continuously monitors the environment to identify potential dangers and fulfils its designated task by avoiding harmful conduct as designed during its training. When the possibility of harm arises, the vehicle should react to avoid it, minimise the damage, or choose the option that results in the minimum harm. In such situations, collision avoidance systems should be designed for the conduct that minimises risks. Indeed, fully simultaneous and perfectly balanced life versus life dilemmas, where all probabilities are equal, are likely to be exceedingly rare; instead, conflicts will typically involve legal interests of varying degrees. Furthermore, it could be argued that, had the programmer designed a better system, the dilemma might have been entirely avoidable; for instance, the vehicle might have braked earlier, preventing the dilemma from arising in the first place.

Therefore, contrary to the widespread perspectives in the literature, it can be stated that the occurrence of isolated, pure dilemmas involving intentional offences will be exceedingly rare. Instead, the emphasis should shift towards analysing real-life scenarios predominantly through the perspective of the duty to develop collision avoidance systems to the highest possible standard. In this context, the debates should centre on whether state of the art collision avoidance systems have been adequately designed and implemented. This focus is also logically more consistent. For instance, the scenario of a self-driving vehicle operating in full compliance with traffic rules when a single individual suddenly steps into its path may be considered a permissible risk. However, in a situation where two individuals unexpectedly step into the path of the vehicle and the vehicle made a manoeuvre to avoid the collision, it would be inconsistent to assess the situation as intentional killing.

Nonetheless, the principle that life holds the highest value must remain inviolable. The argument here only emphasises that in real-world conditions, pure dilemmas are likely to be exceedingly rare. Furthermore, the dilemmas attributed to self-driving vehicles are more likely to arise in the future not on the highways, but in situations where AI-driven systems categorise and profile individuals, requiring them to make choices between such categories; for example, deciding among patients awaiting organ transplants. However, in these contexts, different legal constructs needs to be assessed.

At the end of the study, recommendations for *de lege ferenda* have been examined. Due to the challenges associated with determining criminal liability, it has been proposed in the literature that criminal provisions be introduced, stipulating the placement of dangerous products on the market without adequate safety measures as an abstract endangerment offence, with the occurrence of harm serving as an objective condition of punishability (*objektiver Bedingung der Strafbarkeit*). Such harm could include the occurrence of bodily injury or significant property damage. As a highly pragmatic and feasible proposal, this approach partially addresses the challenges arising from fault-based liability and the determination of causation in criminal law, without the application of strict liability. However, certain concerns can be raised regarding this approach. First, ensuring the product's safety measures could turn into a mere box-ticking practice, as criticised throughout this study. This could result in a superficial compliance, focusing on formal adherence rather than genuinely addressing risks and preventing harm. Moreover, while the suggestion effectively addresses the liability of manufacturers and systems classified as products, it does not account for the non-product AI systems that can be rapidly developed, modified, and deployed on the internet under an anonymous identity. Another issue is that the general theoretical criticisms towards the objective conditions of punishability can also be directed to this approach, particularly concerning the restrictive effect on determining which values fall within its scope. For instance, the exclusion of violations of legal interests such as privacy from punishable acts may pose an issue.

Finally, it must be emphasised that determining which activities are to be deemed acceptable despite their inherent risks ultimately constitutes a matter of legal policy. In this context, it is crucial to approach the concept of permissible risk with great sensitivity, as it effectively establishes an area where liability is excluded. In particular, rather than bearing the responsibility and liability for performing a task directly, delegating such tasks, along with their inherent risks, to AI-driven autonomous systems may itself serve as a basis for liability. Accordingly, I disagree with the prevailing tendency in the literature to categorise individuals as entirely passive merely because such tasks are carried out by autonomous systems. However, if, in the future, the majority of tasks are undertaken by autonomous systems by default, and society as a whole adopts this practice, the role of law would shift from being determinative to primarily explanatory. Indeed, the issues evaluated in this study pertain to systems that are not fully autonomous and entirely independent from humans (in the loop). The current systems are still initiated or activated by humans, and as mentioned, their unpredictability is recognisable. However, in a future where such systems are ubiquitous across all domains and form an integral part of the environment into which humans are born, determining liability will be even more challenging. Furthermore, discussions in the legal literature are still framed around evaluating these systems as distinct from humans. On the other hand, in the near future, it is likely that humans will increasingly integrate systems with partially autonomous features into their own behaviours and functions (and even bodies). In such a scenario, determining whether the conduct under assessment of criminal liability originates from a human behaviour, the artificial autonomous system, or a combination of both will become even more challenging.