Chapter 4: Criminal Liability of the Persons Behind the Machine

The study has so far has focused on the specific challenges posed by criminal offences involving AI-driven autonomous systems, the occurrence of criminal incidents and the doctrinal perspectives on various liability models. In accordance with views suggesting that the existing criminal law framework is insufficient to adequately address these challenges -potentially leading to a “liability gap”- alternative liability models discussed in legal literature have been evaluated. In this context, the study has considered the notion of holding the robot itself liable, non-fault-based liability models, and other proposed frameworks. However, in comparing fundamental differences to criminal liability, it has been concluded that none of these approaches can fulfil the fundamental functions of criminal law, in particular the notion of retributive justice. It should be emphasised that the perception that the criminal liability of the person behind the robot arises because the robot itself cannot bear liability, is not entirely accurate. Criminal law assigns liability to culpable individuals who fulfil the elements of an offence, not merely because another is exempt from liability.

This chapter constitutes the central focus of the study. Accordingly, the liability of individuals behind the machine will be examined within the framework of established criminal law doctrine. Negligent liability will be addressed with particular focus on the *ex ante* and *ex post* challenges posed by AI-driven autonomous systems. Additionally, concepts such as *permissible risk* and the *principle of reliance* will be explored, alongside an examination of dilemma problems that are widely debated in academic literature.

The *actus reus* for the liability of persons behind the machine will not be analysed, as it constitutes a distinct and extensive subject of inquiry that would exceed the scope of this study. Nonetheless the matter would require a particularly nuanced approach, especially in relation to offences committed through omission (*unechtes Unterlassungsdelikt*). As discussed in the the German Federal Court of Justice’s (BGH) *Lederspray* decision, activities such as producing or programming an AI system initially appear as active behaviours, while the failure to continuously update the software or recall a product in the event of a malfunction could be evaluated as omissions[744]. Moreover, in the age of automation, tasks traditionally performed actively by humans are increasingly being replaced by machines. This shift raises the possibility that active and passive duties may interchange[745]. It has been further argued that solely because rules are programmed into the system, such behaviour may not be considered active. The distinction between action and omission remains a matter of judgement and is increasingly evolving with these systems[746].

In a system where AI increasingly takes over human tasks (a trend expected to grow in the future) individuals will engage in fewer active behaviours. Their primary active conduct may be limited to the initial setup of the machine, with subsequent discussions focusing on their guarantor positions. In this context, the debate surrounding whether offences committed through omission are *numerus clausus* holds significant importance from a criminal policy perspective. Particularly in such scenarios, the distinction between action and omission is often deemed irrelevant; the critical question is whether the conduct constitutes a breach of duty and is therefore negligent. In this regard, if the duty of care is taken as the basis without separately evaluating active and omissive behaviour, and if guarantor duties are regarded as equivalent to duties of care, the question of why both companies and individual employees are subject to certain duties of care can be more easily addressed without focusing on guarantor positions or omissions[747].

A. Causality

1. General Challenges with the Causal Nexus for Autonomous Systems

In a Newtonian universe, where determinism is the prevailing paradigm, an understanding of cause-and-effect relationships, as well as their foreseeability, depends on obtaining more information. As more details about events and phenomena are obtained, the probability of a particular outcome can be more accurately calculated. However, as autonomous systems are involved in the causal nexus and interact with the environment, their

[744] See: Chapter 3, Section C(1)(d)(6): "Criminal Product Liability".
[745] LOTHAR, Der Handlungsspielraum, 1974, p. 140, 79 fn. 105.
[746] FELDLE, Notstandsalgorithmen, 2018, p. 250
[747] IBOLD, Künstliche Intelligenz und Strafrecht, 2024, pp. 299-301.

conduct becomes increasingly complex. Consequently, linear causation is increasingly challenged and surprises become inevitable[748].

In cases of systems that are automated rather than autonomous, human behaviour will be directly identifiable within the causal nexus. However, with the emergence of highly advanced "intelligent systems" in the future, the influence of human behaviour on causality concerning outcomes is expected to weaken, thereby raising complex questions of accountability[749]. As AI-driven systems become more autonomous, attributing their generated outputs to the programmer (or the person behind the machine in the given scenario) becomes increasingly challenging. This difficulty grows further, particularly when programming errors are evaluated through the lens of the usual course of events and life experience: criteria that may prove inadequate for addressing the complexities of adaptive systems, which are approaching the limits of such conventional assessments[750]. Besides, the system's autonomous decision-making may interrupt[751] the traditional imputation of liability, making it difficult to directly connect specific actions or failures to the resulting injury[752].

To illustrate, an incident involving Google's chatbot, Gemini, is worthy of note. A student, seeking assistance with their homework, received a disturbing response from Gemini, which included statements such as "You are a stain on the universe" and "Please die". Google has acknowledged the incident, attributing it to the unpredictable nature of LLMs, and stated that measures have been implemented to prevent similar incidents[753]. In this example, pinpointing the exact cause of the chatbot's harmful response -such as inadequate training data, lack of robust safety mechanisms, misinterpretation of user input, testing gaps, or similar factors- is nearly impossible. Furthermore, it is not feasible to attribute this outcome to the actions of a specific individual within a clear cause-and-effect relationship. On the

[748] KARNOW, The application, 2016, pp. 73-74.

[749] JOERDEN, Zur strafrechtlichen, 2020, p. 296 f.

[750] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 177.

[751] The term "interrupt" is not used in the sense of intercepting the causal chain. According to the *conditio sine qua non* formula, rather than the causal link being severed, it is possible for other causal series to contribute to the outcome.

[752] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 445 Rn. 25; ALBRECHT, Fährt der Fahrer oder das System, 2005, p. 375.

[753] VIGILIAROLO Brandon, "Google Gemini tells grad student to 'please die' while helping with his homework", 15.11.2024, https://www.theregister.com/2024/11/15/google_gemini_prompt_bad_response/. For the whole conversation: https://gemini.google.com/share/6d141b742a13. (accessed on 01.08.2025).

other hand, although the precise parameters remain indeterminate, the underlying causes of the recent incident involving Twitter (X)'s AI chatbot, Grok -where it issued insults and threats to numerous users over the course of several days- can nonetheless be generally identified. The main cause of this incident has been attributed to a prompt introduced in early July 2025, following Elon Musk's instruction that Grok should be made "less woke", which led the chatbot to "not shy away from making claims which are politically incorrect"[754].

The role of an autonomous system within the causal nexus might be evaluated as akin to the actions of a third party in the causal relationship between the operator's behaviour and the ultimate outcome. However, this view is not accurate, as AI-driven systems cannot commit acts in the sense of criminal law[755]. As such systems currently lack the capacity to form their own will, liability remains with the person behind the machine. However, this may change if intelligent agents capable of genuine "learning" and memory emerge in the future[756].

In the future, a major challenge concerning AI-driven autonomous systems will arise from scenarios where third parties, such as users, further develop or train the system after their release. Imposing an obligation on manufacturers to oversee all such modifications would be nearly impossible in practice and might hinder the further development and adaptation of AI-driven systems. This raises significant issues regarding causation, in particular attributability to the manufacturer and the limits of the duty of care, including whether manufacturers must anticipate and prevent user errors. For prior chain actors, the issue typically lies in their significant temporal and locational distance from the occurrence of the event, as it happens after their involvement has concluded. Consequently, establishing a causal link becomes challenging[757]. An example of this can be demonstrated in OpenAI's release of ChatGPT's API to third parties, enabling them to further develop and customise the product. Such cases involve numerous

[754] CHAYKA Kyle, "How Elon Musk's Chatbot Turned Evil", 16.07.2025, https://www.newyorker.com/newsletter/the-daily/how-elon-musks-chatbot-turned-evil. (accessed on 01.08.2025).

[755] SEHER, Intelligent agents, 2016, p. 54.

[756] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 588.

[757] GIANNINI/KWIK, Negligence Failures, 2023, p. 58; GOGARTY/HAGGER, The Laws of Man over Vehicles Unmanned, 2008, p. 73 f.; Singapore, Report on Criminal Liability, 2021, p. 14, [para. 2.4].

challenges, including those discussed under the "problem of many hands" which will be addressed below[758].

The growing interconnectedness and complexity of industrial systems increasingly obscure tracing of causal relationships, significantly complicating the determination of liability[759]. The judiciary will need to reevaluate the concept of causation, particularly in cases involving AI-driven systems that behave in ways unforeseeable by their designers or users[760]. Establishing a causal link, especially in the context of product liability, presents significant challenges, including proving the product's harmful outcome indisputably. In such instances, courts may accept causation without demanding scientific certainty, as long as there are no substantial doubts[761]. For instance, this has led to the emergence of the presumption of causality in civil liability as discussed above[762].

Cases where multiple causes contribute to a harmful outcome can be particularly challenging. For instance, in a semi-autonomous vehicle accident, the crash might result from both the vehicle's software incorrectly classifying an object and the driver failing to keep their hands on the steering wheel. If it can be determined that the accident would have occurred even if the driver had kept their hands on the wheel, liability cannot be attributed to the driver. This is because liability requires that the harmful outcome result directly from the specific breach of duty. If it arises from another cause, criminal liability will not be in question[763]. However, the obligation to keep hands on the steering wheel exists to ensure intervention in the event of a potential hazard. Such hazards may also arise from a probable malfunction of the vehicle, and the driver can be obliged to prevent such harmful outcomes within their capacity. If the semi-autonomous vehicle provides the driver with sufficient time to intervene, but the driver fails to act due to not keeping their hands on the wheel, liability would not rest with the manufacturer. In such cases, the driver's breach of duty of care would take precedence in the causal chain[764]. An illustrative example is the

[758] See: Chapter 4, Section D(1): "The Concept of "the Problem of Many Hands"".

[759] HÖTITZSCH, Juristische Herausforderungen, 2015, p. 81.

[760] CALO, Robots in American Law, 2016, p. 23.

[761] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 172 ff.

[762] GRAHAM/THANGAVEL/MARTIN, Navigating AI-Lien Terrain, 2024, p. 201 f.

[763] ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 273.

[764] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 178.

2016 Tesla accident referenced above, in which the driver, despite explicit instructions, became distracted[765].

2. Legal Theories of Causality: Implications for AI-Driven Autonomous Systems

a. Assessment Based on Causality Theories

Issues of negligence and foreseeability are deeply intertwined with notions of causation in legal theory. Concepts beyond the *condition theory* do not limit themselves to examining causation purely from a natural sciences perspective but assess it through the lens of certain values. Before exploring the core issues of negligent liability, which form the backbone of this study, it is essential to briefly highlight the aspects related to causation. Examining whether the differing notions on causation lead to divergent outcomes will also contribute to debates concerning the legal nature of permissible risk. At the core of all these discussions lies the question: can a causal nexus be identified where the person behind the machine's liability can be retrospectively assessed for the harmful outcome in which the AI-driven autonomous system is involved?

Causality is not treated as a fixed scientific concept but is examined differently across disciplines based on their specific needs[766]. If the outcomes of a particular behaviour could be determined with absolute certainty beforehand, assessing the actor's ability to foresee such results would be much simpler. However, aside from the challenges with autonomous systems where the actor's control is increasingly diminishing, the world is already filled with atypical situations and *black swans*[767]. Moreover, some causal relationships are probabilistic; certain behaviours lead to outcomes only in some instances, with varying degrees of likelihood. Besides, although deterministic causality -with its fixed cause-effect relationship- could simplify matters; its application is constrained by the current limits of human knowledge and capacity. Given these limitations and the possibility of alternative causes, courts need to rely on practical "real-world" certainty rather

[765] See: Chapter 2, Section C: "Prominent Cases Highlighting AI-Related Liability".

[766] HILGENDORF, Wozu Brauchen Wir, 2004, pp. 36-41.

[767] TALEB Nassim Nicholas, The Black Swan: The Impact of the Highly Improbable: The Impact of the Highly Improbable, 2nd ed., Random House Publishing Group, 2010.

than absolute proof when faced with alternative explanations[768]. Ultimately, law operates within a framework of constructed fictions.

In the context of precisely identifying the cause-and-effect relationships, the type of causality notion adopted in law becomes crucial. First, the suitability of the *condition theory*, which has its roots in the natural sciences, can be briefly evaluated. Under this theory, the *conditio sine qua non* formula is applied to determine whether a specific act or omission was a necessary condition for an outcome. Accordingly, an act qualifies as a cause of a result if the result would not have occurred but for that act. In fact, multiple factors can contribute minimally or significantly to an outcome in a causal relationship. Therefore, under this theory, the notion of severing the causal nexus becomes inaccurate. Instead, it is only possible for a new causal chain to begin or another causal chain to take precedence, independently producing the outcome[769].

Condition theory is currently the prevalent theory, particularly in German jurisprudence[770]. It maintains objectivity in legal causation by treating all contributing conditions equally without introducing value judgments, while other theories are criticised for incorporating subjective evaluations that undermine scientific neutrality[771]. However, when it comes to AI-driven autonomous systems, thousands of separate conducts performed by hundreds of people involved in the development of AI can ultimately result in unwanted outcomes. As a result, examining causation between these countless actions and the resulting harm significantly complicates the analysis. Consequently, the critique of condition theory lies in its overly broad attribution of causality, leading to absurd results by treating all conditions as equally significant. Hence, to address this, normative criteria are introduced under the framework of *objective imputation* to assess the relevance of causal connections[772].

The *Objective imputation theory*[773] is so-named because it can exclude imputation within the framework of the objective elements of the offence,

[768] HILGENDORF, Wozu Brauchen Wir, 2004, pp. 36-41.

[769] KAUFMANN, Objektive Zurechnung, 1985, p. 269; ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 362.

[770] See: HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 49 Rn. 25.

[771] KÜHL, Wer einen Menschen töte, 2009, p. 325.

[772] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 51 Rn. 33; RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 75 Rn. 7

[773] The term "*objective imputation*" in English has been adopted in this study to correspond to "*objektive Zurechnung*" in German legal doctrine. For the same usage, see:

regardless of the perpetrator's personal circumstances[774]. It should be noted that the concept of objective imputation is not a theory of causality. Rather, causality is first established using the condition theory, and only then is it evaluated whether the objective elements of the offense may be negated based on principles of imputation[775].

The objective imputation theory has evolved significantly since *Honig*'s original formulation[776], and *Roxin* is regarded as the re-founder of the theory[777]. According to the theory, a factual outcome is only attributable to the perpetrator if a legally relevant and disapproved risk that they created materialises in the factual outcome[778]. The risk associated with the use of AI-driven autonomous systems in a specific task, whether it increases or decreases, is particularly significant in this context. In the examination of legally relevant and disapproved risk[779], the focus is not on the overall assessment of the act but rather on whether the perpetrator has taken a fundamentally unlawful risk regarding the outcome. In this context, even someone acting in self-defence can create a legally disapproved risk. The key point is that the risk created by the perpetrator must materialise in the outcome, and it should not be a completely different risk arising from general life hazards, coincidental factors, or independent actions by others that eliminate those of the perpetrator[780].

Accordingly, the creation of a legally disapproved risk means violating a behavioural norm, whether it is written, like the traffic rules in the Road

CHIESA, Comparative Criminal Law, 2014, p. 1096; STUCKENBERG, Causation, 2014, p. 487; ZHAO, Principle of Criminal Imputation, 2024, p. 71.
Some scholars in legal literature prefer "objective attribution" to describe the concept, see: WEIGEND, Germany, 2011, p. 268.
Finally, some scholars use both to correspond the concept, see: DÍEZ/CHIESA, Spain, 2011, p. 506.

[774] GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 159, Rn. 88.

[775] ZIESCHANG, Strafrecht AT, 2023, p. 34 Rn. 84.

[776] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 398 f.

[777] HILGENDORF, Wozu Brauchen Wir, 2004, p. 43.

[778] KAUFMANN, Objektive Zurechnung, 1985, p. 254; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 258; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 56 Rn. 46.

[779] The terms "legally disapproved risk," "legally impermissible," "legally relevant danger," and "legally prohibited conduct" are all used interchangeably, with no substantive difference between them. See: KÜHL, Strafrecht AT, 2017, p. 43 f. Rn. 43.

[780] RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 85 Rn. 48 f.

Traffic Act (StVO)[781], or unwritten, like the rules of medical practice. For instance, a driver who entirely complies with the StVO operates within a permissible risk, and a resulting death cannot be objectively attributed to them, despite the presence of causation, as no legally disapproved risk was created[782]. On the other hand, in cases where the risk is reduced, the outcome cannot be objectively attributed to the perpetrator. However, this differs from cases where the risk is altered, thereby establishing a new, independent risk that materialised in the outcome. Such cases are objectively imputable, although criminal liability may be excluded on other grounds[783]. If the outcome resulted not from the initial risk created by the perpetrator but from the materialisation of a different risk, the result cannot be objectively imputed to the perpetrator[784].

According to proponents, the objective imputation theory applies to both intentional and negligent crimes, but is most impactful in cases of negligence. Accordingly, there is no lack of due care if the perpetrator has not created any legally relevant risk from the outset. Furthermore, negligence is not merely the omission of due care but involves creating a risk that exceeds permissible limits, falls within the protective purpose of the offence, and materialises in an outcome defined by the legal elements of the crime[785].

The objective imputation theory has faced criticism in literature from various perspectives. First, although it was introduced to limit the scope of the objective elements of the crime and the broad extent of the *conditio sine qua non* formula; no precise content or consensus on its practical application has been achieved, despite significant efforts. On the contrary, its use has been reduced to an appeal to common sense notions of right and wrong in many cases[786] and to subjective value judgments rather than precise

[781] Straßenverkehrs-Ordnung (StVO), enacted on 06.03.2013, last amended on 11.12.2024, https://www.gesetze-im-internet.de/stvo_2013/BJNR036710013.html. (accessed on 01.08.2025).

[782] KÜHL, Wer einen Menschen töte, 2009, p. 326.
According to the theory, the standard of care is determined objectively and *ex ante*, considering any special knowledge of the perpetrator. If a diligent third party cannot recognise the risks *ex ante*, such risks are disregarded. See: WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 822, Rn. 90

[783] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 293 f.

[784] RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 89 Rn. 60.

[785] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1186 f. Rn. 10 ff.; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1126.

[786] HILGENDORF, Wozu Brauchen Wir, 2004, p. 44.

legal reasoning[787]. Furthermore, the theory has been criticised for being frequently employed as a theoretical repository for unresolved problems of elements of the crime and justification[788]. Thus, it has been likened to an *octopus with countless tentacles*, encompassing a wide range of ontologically and normatively heterogeneous areas[789].

Moreover, the theory has been criticised for its misleading claim of being "value-free" as even the basic causality test inherently involves value judgments[790]. It has further been argued that the theory's attempt to explain the unlawfulness of a legal value violation by relying on the unlawfulness itself creates a circular reasoning[791]. Additionally, the theory's reliance on the condition theory to establish a connection between human behaviour and the objective elements of the offence is considered to be logically flawed[792].

As indicated, the concept of risk holds particular importance in the context of this study. In the theory of objective imputation, however, risk itself holds little independent significance since any behaviour causing a result is inherently risky, so the emphasis is on whether the risky behaviour is unlawful or not. However, relying on unclear and broad concepts of risk creation and realisation lacks a convincing principle to limit criminal law aimed at protecting legal interests[793]. Indeed, the concepts of creating a legally relevant risk (exceeding the permissible level) and its realisation in a specific outcome not only embed causal implications themselves[794], but also, they are overly vague; often serving mainly as a flexible justification for intuitively perceived correct results[795].

Criticism of the objective imputation theory extends beyond its vagueness; it is also argued that the cases it seeks to address could be resolved adequately using existing legal principles, making the theory practically

[787] ZIESCHANG, Strafrecht AT, 2023, p. 34 Rn. 86.

[788] HILGENDORF, Wozu Brauchen Wir, 2004, p. 43 f.

[789] SCHÜNEMANN, Über die objektive Zurechnung, 1999, p. 207.

[790] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 56 Rn. 47.

[791] GÖSSEL, Objektive Zurechnung, 2015, p. 22 ff.

[792] *Ibid.*

[793] KINDHÄUSER/HILGENDORF, Vorbemerkung zu § 13 - Strafgesetzbuch, 2022, p. 113 Rn. 103; SCHÖMIG, Gefahren und Risiken, 2023, p. 81.

[794] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 56 Rn. 47.

[795] HILGENDORF, Wozu Brauchen Wir, 2004, p. 35; HILGENDORF, Gefahr und Risiko, 2020, p. 14.

unnecessary[796]. Nevertheless, despite it all, it has been argued that the idea of objective imputation provides more predictable answers[797].

Although not applied in criminal law, the *adequacy theory* (*Adäquanztheorie*), which prevails in civil law, and the *relevance theory* (*Relevanztheorie*) have nonetheless contributed to the development of the objective imputation theory[798]. The adequacy theory aims to break the infinite chain of causation of the *condicio sine qua non* formula into manageable pieces. Accordingly, not every condition is regarded as a cause; but only those based on experience capable of bringing about the outcome are. Atypical causal processes contradicting general life experience and unforeseeable events are thus excluded, ensuring that criminal liability does not extend beyond the capacity of humans to control and manage causal processes[799].

Under the objective imputation theory, objective foreseeability[800] exists if the causal course can be expected based on life experience and the initial danger materialised in the outcome. However, attribution is excluded if the causal nexus was so unusual and improbable that it could not reasonably have been foreseen[801]. Moreover, when determining a causal connection, not all relationships can be deemed deterministic: statistical correlations also exist, where a cause does not consistently lead to the same outcome in all cases. In such instances, past experiences and empirical data determine the likelihood of the outcome. However, with emerging technologies, the lack of sufficient empirical data can lead to challenges, leaving only assumptions, rather than scientific expectations to be made *ex ante*[802]. Indeed, the criteria of life experience in determining causation is ambiguous when applied to AI-driven systems, which continuously reveal new features. For instance, until the *Tay* incident, it could not be considered part of general life experience that chatbots might behave in such a manner; or until the *Aschaffenburg* case, that installation of lane-keeping systems could lead to fatal outcomes if a driver becomes incapacitated, even if some individuals

[796] ZIESCHANG, Strafrecht AT, 2023, p. 34 Rn. 85.

[797] RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 84 Rn. 44.

[798] GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 159, Rn. 86; RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 75 Rn. 8.

[799] STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011., p. 79 Rn. 21; GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 156, Rn. 79 ff; RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 76 Rn. 9.

[800] Foreseeability will be evaluated in detailed below. See: Chapter 4, Section C(4)(a): "The Boundaries of Foreseeability".

[801] RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 90 f. Rn. 62-65.

[802] HILGENDORF, Gefahr und Risiko, 2020, p. 18.

may have anticipated such possibilities. Does this mean that causation should be denied in these cases?

The *conditio-sine-qua-non* formula is a useful tool for identifying the causal nexus; but does not suffice as a comprehensive definition of causality[803]. The doctrine of *lawful conditions*[804] (*Lehre von der gesetzmäßigen Bedingung*) also assumes the equivalence of all factors but avoids hypothetical elimination, by replacing the overall conclusion of condition theory with a detailed chain of lawful conditions; asserting that an action is causal if subsequent external changes -lawfully connected to the action- occur and meet the legal criteria[805]. Thus, it determines causality by assessing whether a connection between an action and its outcome can be explained by known natural laws, addressing some limitations of the condition theory. Although the theory of lawful condition offers a better and more precise method by largely avoiding uncertain hypothetical considerations, it has been argued that it rarely leads to different results and still faces limitations when necessary empirical knowledge is lacking, requiring clarifying discussion in problematic cases[806]. Still, it is suggested that causality problems associated with collective decisions, which are significant in the context of the many hands problem[807], can be addressed by applying the doctrine of the lawful condition[808].

Neither the German Criminal Code (StGB) nor the Turkish Penal Code (TPC) provides a specific explanation regarding causality; leaving the matter to science and jurisprudence. Currently, the condition theory (adopted particularly in court decisions and by part of the doctrine) and the doctrine of lawful conditions are widely recognised in criminal law. The objective imputation theory, on the other hand, has not been applied much in either Turkish or German courts[809]. Yet, they differ not so much in their results as in the nature of their reasoning, as they typically lead to the same practical outcomes[810].

[803] HILGENDORF, Wozu Brauchen Wir, 2004, p. 36.

[804] The English term has thus been adopted. See: STUCKENBERG, Causation, 2014, p. 474.

[805] GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 155, Rn. 74.

[806] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 249; RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 76 Rn. 12.

[807] See: Chapter 4, Section D(1): "The Concept of "the Problem of Many Hands"".

[808] HILGENDORF, Fragen der Kausalität, 1994, p. 566.

[809] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 260; HILGENDORF, Wozu Brauchen Wir, 2004, p. 43; DEMIREL, Taksir, 2024, p. 409 f.

[810] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 225, 235.

b. Distinctive Challenges with Causality

A considerable number of atypical outcomes may arise in the context of AI-driven autonomous systems. However, the explanations provided thus far regarding causality have not sufficiently addressed the matter. Indeed, the unpredictability of such atypical results and the inability to prevent them present distinct challenges. Particularly in the context of negligent crimes, it is theoretically significant whether atypical causal course should be examined under the objective foreseeability[811]. According to one view, even if the act constitutes a necessary condition for the outcome, causation cannot be established if the outcome could not have been foreseen as a typical consequence of the act based on the most advanced scientific and technological knowledge available[812].

The widely accepted principle in contemporary jurisprudence is that, as a rule, the general foreseeability of the outcome is sufficient, while the specific details of the causal nexus leading to that precise outcome are not decisive. An exception arises only when the causal sequence is so far separated from all life experience that it could not have reasonably been anticipated[813]. In the evaluation of risk, entirely improbable occurrences of harm are excluded either through the concept of an atypical causal course under objective imputation or by treating objective foreseeability as a prerequisite for establishing objective negligence[814]. For example, a self-driving vehicle causing an accident due to misperceiving its surroundings through its sensors is (given today's level of knowledge) a probable outcome, whereas its software hacking an information system is improbable. However, if the perpetrator possesses specific knowledge, this must also be taken into account in the *ex ante* objective foreseeability assessment[815].

Another issue that may complicate the causality analysis in offences involving AI-driven autonomous systems is the involvement of a third party's contribution to the causal nexus, which may ultimately lead to an atypical causal process. Undoubtedly, if the involvement is known or objectively foreseeable, it requires a separate consideration. It is argued that particu-

[811] KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 227 Rn. 41.

[812] TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 153.

[813] Federal Court of Justice (BGH), judgment of 12.02.1992, Case No. 3 StR 481/91, reported in NStZ 1992, p. 335. RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 92 Rn. 70.

[814] SCHÖMIG, Gefahren und Risiken, 2023, p. 161.

[815] RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 93 Rn. 74.

larly basic negligent misconduct still reflects a risk that must be anticipated and, therefore, falls within the perpetrator's sphere of responsibility[816]. Examples of this include users' false or misuse of AI-driven systems, lack of proper oversight, manipulation, and similar actions[817]. To illustrate, in a case where a person is intentionally injured, but dies due to the intervening doctor's negligence; if the doctor's negligence does not reach the level of gross negligence, the initial perpetrator remains liable. However, when a third party's misconduct reaches the level of gross negligence, it becomes the predominant factor in the outcome[818]. If both the initial perpetrator and the intervening doctor are roughly equally negligent and contributed to the outcome, both may be held liable for negligent homicide[819]. Nevertheless, even if the perpetrator has created an unlawful risk, the resulting harm cannot be attributed to them if it arose from a distinct risk that was not created by the perpetrator, but by a third party[820].

Another significant challenge with causality is in determining whether the harmful outcome would have still occurred even if the alternative lawful conduct has been followed. For instance, if manufacturer fails to take necessary precautions, such as conducting sufficient tests or carefully selecting training data before releasing an AI system on the market, and the system causes harm due to the insufficient tests, the manufacturer can be held liable for negligence. The key question here is whether the harmful consequence would have occurred with sufficient tests and the proper dataset utilised. This determination is particularly complex, and often nearly impossible, in the context of autonomous systems, largely due to the difficulty of identifying the precise cause of the harm, as elaborated in the *ex post* analysis. Nonetheless, if it can be proven that the harm would have still occurred, the manufacturer cannot be held liable[821]. This is similar to the commonly referred example: if someone is driving at excessive speed and a pedestrian is struck, where the injury could not have been avoidable

[816] *Ibid*, p. 98 Rn. 94 f.

[817] The topic will be examined in detail below within the framework of extending the principle of reliance to machines and exploring whether machines should rely on humans. See: Chapter 4, Section D(2)(c)(2): "Should Autonomous Systems Rely on Humans?".

[818] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 525 Rn. 143.

[819] *Ibid.*

[820] FRISTER, 10. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 133 Rn. 22; ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 524 Rn. 142.

[821] SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 501.

even at the prescribed speed, negligence is excluded due to the lack of realisation of the risk[822].

Finally, as will be addressed particularly under the problem of many hands, issues of cumulative causality may arise. For example, an accident may occur due to an issue stemming from the interaction between two different autonomous systems. However, such cases do not generate a distinct debate beyond the existing ones on cumulative causality and must be resolved on a case-by-case basis. Moreover, atypical causality issues may also occur. These either do not present unique challenges specific to AI-driven autonomous systems and will therefore not be examined further.

B. Intentional Liability

Autonomous systems driven by AI do not exhibit any distinctive characteristics with respect to intentionally committed crimes. Despite the risks associated with autonomy and *black box* issues, if it is possible to determine *ex post* why the crime occurred, the perpetrator can be held directly liable for intentional behaviour. To illustrate, if an individual intends to kill someone using a defective drug, they do not necessarily need to understand the precise mechanism by which the drug produces its effects, similar to AI-driven systems[823].

As highlighted in the 2023 *Global Terrorism Index*, terrorists employ unmanned aerial vehicles (drones) and other AI-driven systems to achieve their objectives[824]. Similarly, through the use of AI-driven systems, it is possible to carry out learning-based cyber-attacks or highly tailored phish-

[822] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1187 Rn. 13.
However, this issue will be examined in greater detail below, with particular consideration given to the enhancement of risk theory (*Risikoerhöhungstheorie*). See: Chapter 4, Section C(5)(b)(3)(b): “Risk Enhancement through Task Delegation to AI-Driven Autonomous Systems: A Legal Analysis”.

[823] SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 448 ff.

[824] LIANG Christina Schori, “Terrorist Digitalis: Preventing Terrorists from Using Emerging Technologies”, Institute for Economics & Peace. Global Terrorism Index 2023: Measuring the Impact of Terrorism, Sydney, March 2023, http://visionofhumanity.org/resources, p. 72. (accessed on 01.08.2025).
For a further example of a target being struck using autonomous drones, see: COTOVIO Vasco/SEBASTIAN Clare/GOODWIN Allegra, "Ukraine’s AI-enabled drones are trying to disrupt Russia’s energy industry. So far, it’s working", 02.04.2024, https://edition.cnn.com/2024/04/01/energy/ukrainian-drones-disrupting-russian-energy-industry-intl-cmd/index.html. (accessed on 01.08.2025).

ing attacks[825]. However, the essential aspect to emphasise in this context is not their remote-control functionality but rather the utilisation of their autonomous capabilities, which holds particular significance in this discussion. For instance, when a basic automated bot is programmed to perform a specific task in a predetermined manner, the focus is not on the system's autonomy. However, if a command is given to accomplish a task and the bot determines how to execute it using its adaptive capabilities, it can then be classified as an autonomous system, raising complex and challenging issues within the scope of this discussion[826]. Conversely, a deterministic system operating on simple if-then rules would be no different from a screwdriver in terms of its functionality.

Regardless of the level of autonomy exhibited by an AI system, if it is deliberately utilised, such as by employing a self-driving vehicle to run over cyclists or deploying a drone to harm civilians, there is no significant challenge in establishing the causal link, and the elements of the crime. In such cases, the AI-driven system functions as an instrument in the commission of an intentional crime[827]. This can be resembled to a scenario where a dog owner directs the animal to attack someone[828]. The key point here is that the person behind the machine must be able to generally know and desire the consequences of their actions. Although it may not qualify as an autonomous robot in today's sense, in a case in the United States, the California Supreme Court stated in *People v. Davis* that, "[i]nstruments other than traditional burglary tools certainly can be used to commit the offense of burglary (...) a robot could be used to enter the building", "... whether that instrument is a hook or a robot"[829].

Intentional crimes were initially considered to constitute exceptional cases in the context of AI-driven autonomous systems[830]. Because the person behind the machine -particularly manufacturers- would very rarely act with deliberate aims, incidents would generally require assessing negligent

[825] MAHMUD, Application and Criminalization, 2023, pp. 7-8.

[826] The implications of an autonomous system causing crimes different from those intended or foreseen have been examined above under the section titled The Natural Probable Consequences. See: Chapter 4, Section C(3): "The Natural Probable Consequence Liability Model".

[827] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 580.

[828] MITSCH, Roboter und Notwehr, 2020, p. 369.

[829] People v. Davis, 18 Cal. 4th 712, 958 P.2d 1083, 76 Cal. Rptr. 2d 770 (1998), https://law.justia.com/cases/california/supreme-court/4th/18/712.html (accessed on 01.08.2025). TURNER, Regulating AI, 2019, p. 118.

[830] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 7.

liability arising from risks associated with autonomy[831]. However, recent developments indicate that a growing number of fraudulent activities (such as phishing and other cyberattacks) are being perpetrated through AI-driven systems in the digital sphere. This trend suggests that such cases are likely to become the subject of increasing jurisprudence.

According to one opinion, if highly advanced robots are considered human-like beings in the future, they can be assessed in parallel with the case of a person causing another to attack a third party. Consequently, legal concepts such as indirect perpetration, instigation, or complicity may become relevant[832]. In my opinion, when a person intentionally utilises a robot, *i.e.*, sets it in motion, the concept of indirect perpetration cannot be applied, regardless of the degree of autonomy involved[833].

Another example of intentional crimes involving AI-driven autonomous systems can be illustrated as follows: a driver in a semi-autonomous vehicle notices that the vehicle is about to hit a pedestrian. Despite having the opportunity to brake, the driver refrains from doing so upon recognising the pedestrian is an old enemy. In this scenario, the crime of intentional homicide by omission arises, because the driver, being in a guarantor position due to preceding dangerous conduct, deliberately refrains from acting. However, with the advancement of AI in the future, if the law evolves accordingly, the guarantor obligation may arise directly from statutory provisions[834]. It has been argued that passengers in a fully autonomous vehicle will not be considered to be in a guarantor position concerning injured individuals following an accident. This is because their sole role is being transported by the vehicle, without exercising any control over its operation. Consequently, their liability does not extend to a guarantor obligation. For these passengers, only the breach of duties to assist and report according to Section 323(c) of the StGB and Article 98 of the Turkish Penal Code may be relevant[835].

I disagree with the given opinion. If so-called passengers are not in a completely passive situation and possess even limited control over the

[831] VALERIUS, Strafrechtliche Grenzen, 2022, p. 124.

[832] MITSCH, Roboter und Notwehr, 2020, p. 372 f.

[833] The arguments advanced by *Hallevy* and other scholars in support of applying the doctrine of indirect perpetration have been analysed in detail. See: Chapter 3, Section C(2): “Indirect Perpetration”.

[834] KANGAL, Yapay Zeka, 2021, p. 96.

[835] MITSCH, Die Probleme der Kollisionsfälle, 2018, p. 75; KANGAL, Yapay Zeka, 2021, p. 96 f.

system, as well as the ability to intervene, and yet fail to do so, their liability may come into question and should be determined based on the specific circumstances of the case. This is particularly significant given the anticipated future in which many tasks will be automated by delegating to autonomous systems, thereby diminishing human control. For instance, individuals who delegate a task, such as transportation, to a self-driving vehicle also create a certain level of risk. The question of whether delegating tasks to AI-driven systems and the risks inherently associated with performing them manually increases or decreases overall risk, will be explored in greater detail below[836]. Accordingly, from a legal policy view, these individuals should bear an obligation to prevent harmful outcomes arising from the risk they created, depending on the circumstances of the specific case. For example, a person seated in the driver's seat of a vehicle equipped with a steering wheel, accelerator, and brake pedals could be considered capable of intervening. By contrast, in the case of vehicles such as *Tesla*'s recently unveiled *robotaxis*[837], which lack these features, passengers would have no control or means to intervene. Naturally, the law cannot hold individuals responsible for outcomes they have no control over. However, even in this case, particularly from a legal-policy perspective, it should be debated whether the act of actively initiating the journey poses a risk, despite the individual being in a completely passive position during the journey.

Another example can be demonstrated with Google's *Gemini AI*. When *Gemini AI* begins insulting users, a duty to prevent such conduct arises for Google, analogous to the principles examined in product liability cases. Should the company fail to take necessary measures against such malfunctions, particularly in the case that the chatbot will inevitably continue to insult people, and deliberately observe the situation by omitting, intentional liability may come into question (insult is a criminal offence that can be committed intentionally under Article 125 of the Turkish Penal Code and Section 185 of the dStGB)[838]. Yet, determining which individuals within the company would bear liability requires a separate analysis.

[836] See: Chapter 4, Section C(5)(b)(3)(d): "Delegating Tasks to AI-Driven Autonomous Systems: An Alternative Approach for Liability".

[837] TAYLOR Josh, "Elon Musk unveils Tesla Cybercab self-driving robotaxi", 11.10.2024, https://www.theguardian.com/technology/2024/oct/11/elon-musk-unveils-tesla-cybercab-self-driving-robotaxi; https://www.tesla.com/we-robot. (accessed on 01.08.2025).

[838] An opinion on the matter argues that if the manufacturer, after identifying the situation, fails to intervene and take measures; their inaction may constitute partici-

In cases where AI is used as a tool in the commission of crimes, considering that it may amplify the impact of such offenses, it may be appropriate to stipulate it as an aggravating factor of the criminal penalty, due to the convenience and disruptive effect provided by technology[839]. Moreover, it is proposed that crimes committed using AI-driven autonomous systems should classify as "weapons", thereby serving as a factor to increase the punishment[840].

Finally, a report prepared by *Singapore Academy of Law Reform Committee* in 2021 highlights that, in Singapore, existing criminal norms are likely to address various scenarios involving the malicious use of AI. However, it emphasises the uncertainty regarding whether they can adequately cover all potential situations. For instance, it has been stated that intentionally blocking signals to an AI system's sensors and causing it to harm someone, would constitute intentional injury under Section 350 of the Singapore Penal Code[841]. However, concerns have been raised that AI systems could be employed in a variety of harmful actions that may fall outside the scope of existing criminal norms. Furthermore, the classification of AI systems as "weapons" under Articles 324 or 326[842] has also been discussed[843].

C. Negligent Liability

1. The Rationale Behind the Concept of Negligence in Criminal Liability

The foreseeability and avoidability of the consequences of actions, their voluntary nature and the resulting responsibility have been subjects of philosophical and legal debates since the time of *Aristotle*, and even earlier[844]. The question of which behaviours individuals should be condemned or blamed for, and the extent to which such condemnation is appropriate,

pation in the ongoing offences through omission, see: KANGAL, Yapay Zeka, 2021, p. 98.

[839] MÜSLÜM, Artificial Intelligence, 2023, p. 139; ÖZTÜRK, Derin Sahte, 2021, p. 78.

[840] KÖKEN, Yapay Zeka, 2021, p. 267.

[841] Singapore Penal Code 1871, 2020 revised edition, 16.09.1872, https://sso.agc.gov.sg/Act/PC1871?ProvIds=P416-#pr350-. (accessed on 01.08.2025).

[842] Articles 324 - 326 of Singapore Penal Code, https://sso.agc.gov.sg/Act/PC1871?ProvIds=P416-#pr324-. (accessed on 01.08.2025).

[843] Singapore, Report on Criminal Liability, 2021, p. 25 f., [para. 4.6 ff.].

[844] *Aristotle* emphasises a behaviour's voluntariness and its connection to foreseeability when determining liability. Even natural forces like the wind can be foreseeable, and in certain situations, can lead to holding a person liable. See: TAYLOR C. C.

remains a central point of discussion. One significant issue is whether blame should be assessed on the basis of objective criteria or on the subjective state of the perpetrator.

The distinction between culpability and blameworthiness plays a crucial role in legal judgments, particularly in cases of criminal negligence. These cases often involve individuals who did not intend to cause harm but whose lack of due care resulted in harm. Differentiating between these concepts is essential in deciding whether to impose punishment based on moral fault (culpability) or merely on the occurrence of a wrongful act under an individual's control (blameworthiness)[845].

Liability for negligence serves to ensure adherence to generally expected safety standards, promoting the recognition and mitigation of risks[846]. In this context, it can be argued that the primary function of negligent liability is to encourage individuals to act with greater care and diligence. It is not sufficient for a law-abiding individual to avoid outcomes that they deem possible; they must also take measures to recognise potential causes of such outcomes through their behaviour in order to prevent harm[847]. Nevertheless, punishing every instance of carelessness in social life would be neither reasonable nor acceptable. Accordingly, in both German and Turkish legal systems, negligent crimes are regarded as exceptional and are only punishable when explicitly prescribed by law, in contrast to intentional crimes.

2. Advancing Technologies and Negligence

Technological advancements have increasingly brought the various dimensions of negligent liability into focus for deeper analysis and debate. Scientific and technological developments, especially since the beginning of 20th century, resulted in a highly complex and ambiguous evolution in how negligence is assessed. The inherent hazards associated with new technologies have led to a significant increase in negligent acts arising from risk-taking and diminished control, thereby making negligence a central concern in

W., ARISTOTLE Nicomachean Ethics, 2006, Book III, 1109b ff. p. 16 ff., 168, fn. 18; LÜBBE, Erlaubtes Risiko, 1995, p. 951 ff.

[845] BERMAN, Blameworthiness and Culpability, 2024, p. 1.

[846] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 179 f. Rn. 36.

[847] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 403.

criminal law[848]. In this regard, specific provisions to address the negligent endangerment of public safety has been introduced, particularly in cases where such technologies might result in significant risks, such as explosions caused by the release of nuclear energy or explosives[849].

While technologies often simplify and enhance daily life, they can also result in harmful consequences. Traditionally, harmful outcomes resulting from human actions have been addressed under criminal law. However, given the risks posed by machines, liability for negligence may also extend to the person behind the machine. In traditional automated systems, even when it may be difficult to foresee the exact cause of harm, control ultimately remains mainly with humans, and harm can often be prevented through proper design, maintenance and oversight. Negligent liability typically arises from deficiencies in these.

In autonomous systems, on the other hand, control diminishes; but does not vanish entirely. Particularly, manufacturers bear significant control and responsibility in the development and training of AI systems. However, even they cannot fully predict every conduct of their creations, nor can they always pinpoint the precise causes of harmful outcomes when they occur[850]. Examining responsibility in the utilisation of AI systems through the control perspective offers a logical approach. If the manufacturer's control is primarily situated in the design phase, the focus should be on ensuring a robust and safe design. If responsibility relates to adapting the system to new circumstances via software updates, then focus must be directed towards this aspect. Similarly, when users have control over the system, their potential liability must also be considered. The key challenge lies in setting the scope of these responsibilities.

The function of negligent liability in urging individuals to act with greater care is particularly significant in this context. For instance, in the 2015 case of a South Korean woman whose hair became entangled in a robot vacuum cleaner while she was sleeping[851]; the incident highlights the evolving challenges of technology-related liability. At the time, robot vacu-

[848] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 232 f.

[849] SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 257 f.

[850] This issue has been addressed above under the *ex ante* and *ex post* evaluations, See: Chapter 1, Section E: "Distinctive Challenges of Crimes Involving AI-Driven Autonomous Systems".

[851] McCURRY Justin, "South Korean woman's hair 'eaten' by robot vacuum cleaner as she slept", 09.02.2015, https://www.theguardian.com/world/2015/feb/09/south-korean-womans-hair-eaten-by-robot-vacuum-cleaner-as-she-slept. (accessed on 01.08.2025).

um cleaners were still in the early stages of development and widespread adoption, and their mapping of home environments and responses to sensory inputs were relatively underdeveloped. In regions like South Korea, where it is common for people to lie or sleep on the floor, developers might not have foreseen such risks at the time; and it may not have been legally reasonable to expect them to do so (the topic is open for discussion). However, if a similar design flaw were to result in harm today, both civil and criminal negligence liability could be considered. This progression reflects how liability frameworks incentivise manufacturers to adopt more cautious approaches and incorporating these considerations into safer designs.

3. Theoretical Foundations of Negligent Liability in AI-Driven Autonomous Systems

This study does not aim to provide a comprehensive analysis of negligent liability in general, and therefore, will not follow the structure or methodology typically adopted in criminal law textbooks. Instead, it is narrowly focused on criminal liability in cases involving AI-driven autonomous systems. In this context, the analysis will address critical questions, particularly under which circumstances the person behind the machine may be held liable for negligence and the scope of such liability and duty of care. Special attention will be devoted to identifying which risks can reasonably be recognised, averted; or mitigated; the legal expectations that can be imposed on individuals, the foundations of the duty of care, and the principles for determining its standards. This includes an analysis of the appropriate reference point, specifically whose perspective should be adopted in defining these standards.

a. Fundamentals

In the criminal codes of certain jurisdictions, including Germany, negligence is not explicitly defined, leaving its interpretation to legal doctrine and judicial practice. Since the German Criminal Code (StGB) does not provide a definition of negligence, it has been argued in the literature that a degree of ambiguity arises in its application. It is likened to the proverbial "*sword of Damocles*" perpetually hanging over individuals, who, despite their best efforts, may find it impossible to completely refrain from certain types of conduct to avoid liability[852].

In German criminal law, there is a tendency to define negligence in a manner analogous to its conceptualisation in civil law, particularly as a breach of the duty of care pursuant to Section 276(2) of the (BGB)[853]. Although the scope of the duty of care in criminal law closely aligns with the standards applied in civil law, and the requirements of criminal law should not be stricter than those of civil law[854], significant differences exist between the two. Mainly, criminal negligence requires not only an objective breach of the duty of care but also a subjective assessment of whether the harm was foreseeable and avoidable based on the perpetrator's individual knowledge and abilities[855]. Another view, while recognising the need for terminological consistency within the legal system, refers to the German Federal Constitutional Court's decision stating that, in a complex legal system, it is not unusual for legal terms to have different meanings in different areas of law[856]. Hence, it has been argued that the content of negligence in criminal law must differ from that in civil law, as civil law governs relationships between individuals and aims primarily at compensation, whereas criminal law is concerned with punishment[857].

Due to the diversity of concepts surrounding negligence, there is no definition of the term that is fully agreed upon[858]. In this context, negligence is generally defined in literature as the violation of a duty to act carefully and the recognition of the realisation of the elements of the offence[859]; violation of an objective duty of care in the event of objective predictability of the occurrence of the result (for result crimes)[860]; or the unintentional causation of an objectively foreseeable and avoidable unlawful situation

[852] DUTTGE, StGB § 15 MüKo, 2024, Rn. 37.

[853] This aligns with principles already addressed in the objective imputation theory. See: FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 167 Rn. 2.

[854] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 216.

[855] HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 448 f..

[856] Federal Constitutional Court (BVerfG), decision of 18.10.1989, Case No. 1 BvR 1013/89, reported in NJW 1990, p. 241.

[857] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 216; DUTTGE, Zur Bestimmtheit, 2001, p. 233 ff

[858] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1157, Rn. 208.

[859] SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 262 f.

[860] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1101.

through the breach of a duty of care[861]. Another definition states that a person acts negligently if, in light of the circumstances, they create or fail to prevent a foreseeable, avoidable, and legally required avoidance of a situation that leads to an unjustified fulfilment of an offence, given their individual conditions[862].

In criminal law, the examination of negligence is initially based on the foreseeability of the harmful outcome. Conducting a negligence assessment only for foreseeable outcomes prevents liability from becoming limitless and ensures that individuals are not held accountable for results that even the most cautious person could not have anticipated. Some even argue that punishing unconscious negligence breaches the principle of culpability, as it seems unjust to hold someone liable for failing to perceive a situation they were not consciously aware of, which requires a stronger link between actions and mental state[863]. However, the role and position of foreseeability within criminal law analysis varies depending on the perspective adopted[864]. Some views consider objective foreseeability as part of objective imputation, as outcomes that are not objectively foreseeable cannot be objectively attributed[865]; while others examine it within the framework of objective negligence[866].

For instance, in a typical analysis adopting the objective imputation theory, a voluntary act must be established along with causality, objective breach of the duty of care[867], and objective imputation. Within the scope of objective imputation, factors such as objective foreseeability, objective avoidability and the realisation of the result within the protective purpose of the norm are examined. Accordingly, the analysis of objective imputation is crucial in cases of negligence, as the relationship between the breach of duty and the protective purpose of the norm holds particular significance. Additionally, subjective foreseeability and the subjective breach of the duty of care (*i.e.*, the subjective ability to fulfil the duty of care) are assessed

[861] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 555 Rn. 20.

[862] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 195 f. Rn. 87c, 87f.

[863] For the evaluation of this critique, see: FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 168 Rn. 4.

[864] DEMIREL, Taksir, 2024, p. 375- 379.

[865] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 580, Rn. 142.

[866] KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 226 Rn. 36.

[867] Objective breach of duty of care can overlap with the criteria of the creation of a legally disapproved risk within objective imputation.

under the element of guilt (*Schuld*)[868]. Furthermore, it is asserted that the subjective dimension of negligence is rarely problematic in actual cases. As a general principle, it can be presumed that conduct which is objectively contrary to a duty of care and is foreseeable, would also have been subjectively recognisable by the individual. Accordingly, situations such as a lack of intelligence, poor memory, gaps in knowledge, lack of experience, age-related cognitive decline, sudden loss of capacity, or states of shock and confusion do not give rise to the subjective element of negligence[869].

The matters outlined above are also relevant when negligence is analysed through its objective and subjective dimensions within a two-stage evaluation framework. While negligence was traditionally examined under the concept of guilt, the dominant contemporary view endorses a two-stage assessment[870] and that negligence should not be confined solely to an analysis under guilt[871]. Although the matter is theoretically relevant to various aspects; for the purposes of this study, as examined below, its significance lies specifically in determining the concept and boundaries of negligent liability based on whom the standard of care is assessed. For instance, it raises the critical question of whether the liability of an individual developer who creates and releases a generative AI for public use on the internet is equivalent to that of a *Big Tech*[872] company developing a comparable AI system.

According to proponents, negligence has a dual nature; manifesting in both behavioural and guilt forms. In the objective dimension, the issue of whether there has been a breach of an objective duty of care when the outcome was objectively foreseeable is determined. Conversely, the subjective dimension shifts focus to the perpetrator rather than the act itself, as this stage concerns the subjective imputation of wrongdoing. Here, the inquiry examines whether the individual, considering their specific characteristics and abilities, was personally capable of meeting the requirements of the

[868] RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 531 Rn. 12.

[869] The instances of negligent undertaking are reserved.
VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1137, Rn. 158; JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 594; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 550 Rn. 84 ff.

[870] KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 174 Rn. 21.

[871] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 166 Rn. 16.

[872] The term “Big Tech” refers to the highly influential dominant technology companies known for their significant economic, social and cultural impact (such as Alphabet (Google), Amazon, Apple, Meta, Microsoft).

objective duty of care and subjectively foreseeing the occurrence of the harmful outcome[873].

In contrast, according to the individualising theory, which argues that a two-stage analysis of negligence is unnecessary, any legally relevant subjective factors are already considered during the assessment of the breach of the duty of care, making additional deliberation of subjective elements superfluous[874]. Incorporating a subjective element, especially in cases of unconscious negligence, by requiring awareness of risk conditions as a mandatory criterion, is overly restrictive and impractical[875]. This approach individualises negligence within the framework of definitional elements of the offence; examining it through a normative perspective that considers the perpetrator's individual abilities and knowledge as limiting factors[876]. Besides, the two-stage analysis is grounded in the causal theory of action, whereas under the final theory, such an analysis is deemed unnecessary[877].

Despite contrasting views, it has been widely argued that the difference between two perspectives are less significant than the intensity of the debate surrounding it might imply[878]. A key factor in this context is the significant role played by the consideration of special knowledge and abilities[879]. Indeed, apart from some minor differences, there is virtually no practical difference between these two approaches, particularly with

[873] For a detailed assessment, see: WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 619, 1102 f.; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 259 Rn. 7; JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 564; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 232 Rn. 63; VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1136 f., Rn. 154 ff.; ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 177, 180.

[874] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 312 f. Rn. 29 ff.

[875] For an evaluation, see: VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1136 f., Rn. 154.

[876] DEMIREL, Taksir, 2024, p. 388 f.

[877] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 190 f. Rn. 81 f.; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 308 Rn. 58 f.
For the criticisms of two-stage analysis of negligence and the view that it should be positioned solely within the domain of wrongdoing (*Unrecht*), see: MERAKLI, Ceza Hukukunda Kusur, 2017, p. 351.

[878] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1201 Rn. 56.

[879] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 581 Rn. 143.

regard to the principle of permissible risk and reliance[880]. Both of these perspectives agree that for individuals with below-average abilities, criminal liability should not exceed their capacity except in cases of negligent undertaking. The main difference lies in cases of above-average abilities: the individualising theory demands the use of exceptional skills, while the objective theory only requires what is generally expected. However, even this difference is softened, as the two-stage analysis allows special standards for experts and the individualising theory usually aligns with the objective standard of permissible risk and the principle of reliance[881]. Nevertheless, this distinction plays a minor role in practice because courts often infer subjective negligence from objective standards, and those citing below-average abilities face accusations of prior negligence, particularly negligent undertaking[882].

The legal question of what an individual could reasonably have been expected to foresee is further accompanied by the issue of liability for consequences that were actually foreseen. This is particularly relevant in the context of AI-driven autonomous systems, such as self-driving vehicles, where the knowledge of potential risks, including the possibility of traffic accidents, and the gradually emerging statistical data in this area, are of significant importance. While some scholars assert that general considerations of danger are insufficient, arguing instead for the necessity of awareness of a specific risk or probability rather than a mere possibility to establish conscious negligence[883], this view is criticised for creating a gap between conscious and unconscious negligence unless the latter is broadened to cover underestimated risks[884].

Conscious negligence, although not explicitly defined in the StGB, is understood in legal literature as occurring when an individual acts carelessly or engages in impermissible risky behaviour, while recognising the not entirely remote possibility that circumstances fulfilling the elements of a criminal offence may exist or arise. Despite this recognition, the individual

[880] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 181 f., 192 f. Rn. 43 f. 87 f.; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 297, 310 Rn. 18 f., 66.

[881] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1201 f. Rn. 56.

[882] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1137, Rn. 156.

[883] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 568.

[884] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1190, Rn. 289.

seriously, rather than vaguely, trusts that the offence will not occur[885]. In this respect, it differs from unconscious negligence, which arises when an individual fails to consider the possibility that their actions could result in the fulfilling of a criminal offence, thereby failing to recognise the associated risk[886]. Furthermore, the distinction between conscious negligence and *dolus eventualis* is not always easy to delineate[887].

Under German criminal law, in addition to the concept of unconscious negligence, the notion of recklessness (*Leichtfertigkeit*) is also recognised. Recklessness represents an elevated degree of negligence, reflecting greater wrongdoing and culpability. Unlike simple negligence -whether conscious or unconscious- recklessness is required as a prerequisite for liability when specifically mandated by law, as in Sections 239(a)(3), 239(b)(2), and 316(c)(3) of the StGB. Although not explicitly defined in the StGB, recklessness is comparable to gross negligence in civil law but is understood more narrowly in criminal law; with regard to the individual abilities and knowledge of the perpetrator, which are decisive for determining culpability[888]. While not among the typical crimes associated with AI-driven autonomous systems, there is no legal obstacle to applying these provisions to such instances insofar as they align with the nature of the conduct in question. In this context, the explanations concerning recklessness should be considered with respect to the person behind the machine.

b. The Legal Basis of Duty of Care

The theoretical debates surrounding the structure of negligence are fundamentally concerned with the concept of breach of duty of care. However, the question of what constitutes the source of duty of care is particularly

[885] *Ibid*, p. 1189 f., Rn. 287; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 220 Rn. 7; FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 167 Rn. 2.

[886] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1106; JOERDEN, Zur Differenz zwischen Vorsatz und Fahrlässigkeit, 2015, p. 46; FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 162 Rn. 9.

[887] JOERDEN, Zur Differenz zwischen Vorsatz und Fahrlässigkeit, 2015, p. 49 ff.
For an assessment from the perspective of Turkish law, see: AKTAŞ, İnsan Öldürme, 2015, pp. 15-21.

[888] HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 324 Rn. 837; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 294 Rn. 6; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 220 Rn. 10; FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 163 Rn. 12.

significant in the context of emerging and exponentially advancing technologies, such as AI-driven autonomous systems. The logic behind this is clear: each passing day surpasses the expectations of the day before. Companies developing AI allocate substantial resources to these technologies, with significant budgets driving continuous improvement through research and development. To illustrate, the duty of care cannot be assumed to remain unchanged even between the commencement of research for this study and its completion; consequently, an instance which was not regarded as a breach of the duty of care at the beginning might be evaluated as such by the time the study concludes[889]. Similarly, one might question whether a new collision-avoidance system developed by *Tesla* could shape the duty of care applicable to comparable systems developed by *Waymo*. To address such questions, theoretical explanations are provided under this section, and the issue of whether adherence to standards can be considered within the scope of permissible risk will be examined through concrete examples below.

The duty of care may arise from both written and unwritten rules that collectively establish standards of responsible behaviour across various contexts and fields[890]. Written rules, such as statutory provisions, constitute a primary source and are not confined to legal statutes. For instance, beyond traffic laws (e.g., the StVG), technical safety standards and recognised medical protocols explicitly establish obligations to ensure safety and prevent harm. These codified rules are frequently formalised in written form, with their content derived from accumulated professional expertise and societal experience, particularly aimed at addressing risks and recurrent issues[891]. Other written legal rules, such as those governing parental responsibilities

[889] For instance, at the beginning of this study, OpenAI's GPT-3 was accessible to a limited audience, and evaluations were based on their examples of GPT's malfunction. However, these examples were replaced as they were surpassed by more recent ones. As a brief historical note, it is noteworthy that while generative AI was initially considered groundbreaking for producing images such as avocado-shaped chairs, it has now advanced to the point of creating highly realistic videos. By the time this text is read, it is highly probable that even more astonishing capabilities will have emerged, and the creation of such videos may well be regarded as commonplace.

[890] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 299 Rn. 26; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 261 f. Rn. 19 f.

[891] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 557, Rn. 28 ff.; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 531 Rn. 16 f.; AKBULUT, Ceza Hukuku, 2022, p. 502; ZAFER, Ceza Hukuku, 2021, p. 347.

(Section 1626(1) of the BGB or Article 327 of the Turkish Civil Code)[892] further contribute to defining the scope of the duty of care in certain areas[893].

In addition to written rules, unwritten norms also serve as a significant source of duty of care; particularly in areas where official rules are absent or insufficient due to various reasons. These unwritten norms are rooted in shared societal experience, professional practices, and sometimes even common sense[894]. In certain professions, the obligation to act prudently may arise not only from the formal rules governing the profession but also from customary practices and traditions[895]. Additionally, in fields such as hunting, sports, etc. where hazardous activities may occur, the law generally does not prescribe a specific detailed course of behaviour; but imposes general safety regulations and requires the responsible party to observe due diligence. In such situations, general safety principles require individuals to act with due care[896].

Professional and sector-specific standards play a crucial role in further defining the duty of care. Particularly, such written rules may be established not only by official authorities but also by professional organisations, which often develop standards and guidelines based on their expertise and experience to address potential risks. Thus, significant guidance referring to responsible behaviour is also provided by technical regulations, safety guidelines issued by associations or, in medical practice the recognised rules of medical art. What needs to be assessed in this context is whether the guidance is merely advisory in nature[897]. However, although non-legal norms like DIN standards are important in defining diligent behaviour[898], they are generally designed for civil law purposes and serve only as indicators in the context of duty of care for criminal liability[899].

[892] DEMIREL, Taksir, 2024, p. 178.

[893] HEINRICH, Strafrecht AT, 2022, p. 443 Rn. 1010.

[894] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 299 Rn. 26; ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 269 ff.

[895] ZAFER, Ceza Hukuku, 2021, p. 347.

[896] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 239.

[897] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 181 f. Rn. 56; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 299 Rn. 26; KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, p. 202.

[898] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 299 Rn. 26.

[899] BECK, Intelligent Agents and Criminal Law, 2016, p. 139.

To prevent dangers and negligence, comprehensive systems of licensing requirements and regulatory prohibitions, such as those in Germany, are employed. Various legal norms regulate the marketing of hazardous items, technical equipment, food, toys, and pharmaceuticals and other similar things based on their nature. Additionally, civil liability for damages already serves as a significant and often sufficient deterrent against negligent actions[900].

Customs and practices shaped by experience and expertise, even if not yet formalised into written norms, can serve as a source of the duty of care[901]. For example, the training and developing of AI systems must align with the "state of the art" in science and technology[902], as the applicable standards in this field are subject to constant change. In this regard, adhering solely to industry practices may not be sufficient, as such practices often lag behind the *state of the art*. Manufacturers are therefore required to continually update their products to address newly identified risks and to ensure compliance with evolving safety standards and expectations[903]. Moreover, in cases where even the established standards are disregarded during the development of AI systems, the resulting product will inherently contain a design flaw, thereby breaching the duty of care from the moment it is introduced to the market[904].

Consequently, adopting new risk-reducing measures introduced by other AI developers known in the sector is crucial to fulfil the duty of care. This is particularly important in industries (such as self-driving vehicles) where only a few large-scale companies dominate the state of the art due to factors *inter alia*, high costs; making it essential for developers to keep pace with the higher standards set by others. These companies must continually conduct research and development to both improve their products and minimise the risks associated with them. The requirement for one company's developed method to be followed by others could disincentivise inno-

[900] SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 267 ff.

[901] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 557, Rn. 28 ff.; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 531 Rn. 16 f.

[902] In this context, the term 'state of the art' is used to describe the current leading edge of innovation and the most advanced solutions available. On the other hand, while the term 'state of the science' is used to refer to the broader scope of established knowledge, emerging research directions, and underlying theories; 'state of the technology' refers to how these scientific insights are translated into practical, widely implemented tools and processes.

[903] Federal Court of Justice (BGH), judgment of 16.06.2009, Case No. VI ZR 107/08, (Airbag case), reported in NJW 2009, p. 2953 f.

[904] VALERIUS, Strafrechtliche Grenzen, 2022, p. 131.

vation, research and development efforts. It is the responsibility of the legal system to prevent companies from collectively deciding to avoid developing risk-mitigating measures. Yet, even today, vehicles with varying levels of safety and affordability are in the market to accommodate different budgets. This issue will be addressed separately in the context of permissible risk.

In both civil and criminal law, the source of the duty of care may, in some cases, stem not only from contractual or private regulations but also, in addition to the aforementioned ones, from the general principle of refraining from harm when engaging in activities that pose an increased risk to others. This principle is particularly important in the field of robotics, where many aspects and behaviours remain unregulated, and there is a lack of general accumulated experience[905]. In such activities, the unpredictability of AI-driven autonomous systems is, to some extent, anticipated, giving rise to a duty of care.

The question may arise as to whether an operator who, despite recognising that a robot is likely to malfunction, fails to intervene and thereby contributes to a harmful outcome, can be held liable for negligent (or even intentional) conduct. Such a duty to act may stem from a guarantor position established by legal or contractual provisions, or by the creation of a danger. In the field of robotics, a guarantor position may initially arise due to the increased risks associated with the use of such systems[906]. The duty of care should increase proportionally with the likelihood of harm[907]. Still, although risk analysis and increasing knowledge of the circumstances facilitate identifying potential consequences of actions; they cannot serve as the primary indicator for criminal liability. This is because known risks may be ultimately acknowledged, necessitating a distinct evaluation under the permissible risk doctrine[908].

To illustrate the duty of care for a driver in a semi-autonomous vehicle, these obligations may include measures both before and after the vehicle is activated (as specified in the StVO and StVZO[909], *i.e.*, written legal rules). Pre-activation duties include actions such as keeping the software up to date by installing manufacturer-provided updates, adhering to system

[905] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 175.

[906] *Ibid*, p. 179.

[907] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 261 f. Rn. 19 f.

[908] BECK, Intelligent Agents and Criminal Law, 2016, p. 141.

[909] Straßenverkehrs-Zulassungs-Ordnung (StVZO), enacted on 26.04.2012, last amended on 10.06.2024, https://www.gesetze-im-internet.de/stvzo_2012/BJNR067910012.html. (accessed on 01.08.2025).

warnings and familiarising oneself with the system's functionality as well as checking the vehicle's functioning[910]. Post-activation responsibilities may arise from failing to take control when requested as well as failing to override or deactivate the system in cases of obvious malfunctions[911].

To sum up, the duty of care is derived from a multifaceted framework encompassing written legal rules, behavioural standards, professional guidelines, administrative, operational and usage instructions, as well as unwritten norms and, where required, following the *state of the art*[912]. This dynamic interplay ensures that the duty of care remains both comprehensive, dynamic and adaptable to the challenges posed by evolving practices and advancing technologies. In light of the complex and layered sources of the duty of care, lawmakers may in the future impose specific obligations on manufacturers and operators of AI-driven autonomous systems; potentially through checklists or codes of conduct[913]. However, this approach entails a significant risk of reducing the fulfilment of the duty of care to a mere bureaucratic exercise, detached from the practical realities of evaluating risks. A purely formal assessment would fail to genuinely minimise the risks posed by AI-driven autonomous systems in real-world scenarios. Instead, it may function as legal fiction, absolving those behind the machines of true accountability.

Determining the source of the duty of care is essential for defining its scope and boundaries. The lack of clear legal criteria for negligent behaviour creates uncertainty for legal practitioners as well as developers, and raises concerns about compliance and legal certainty which are referred

[910] Just as it is impossible for a human driver to operate a vehicle when the windshield is completely covered with snow or mud, the same logic applies to self-driving vehicles that perceive their environment through sensors. A sensor obstructed by dirt, ice, or as in the 2016 incident, a moth, can impair the vehicle's proper operation and lead to harmful outcomes. Therefore, ensuring the proper functioning of these sensors falls within the responsibilities of the person operating the vehicle. Nevertheless, even if the vehicle operates with a low-level driving assistance feature, the manufacturer fulfils its duty of care by ensuring that the vehicle alerts the driver and requests a complete takeover of control when necessary. MARKER Jason, "Tesla Autopilot disabled by giant moth in Nevada desert", 12.05.2016, https://www.autoblog.com/news/tesla-driver-attacked-by-mothra-in-nevada-desert. (accessed on 01.08.2025).
See also: VALERIUS, Sorgfaltspflichten, 2017, p. 14 f.

[911] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1122; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, pp. 159-164.

[912] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1213 Rn. 96.

[913] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 179.

to in Article 103(2) GG and Section 1 of the StGB[914]. To mitigate such issues, extensive legal debates in advance are crucial for avoiding conflicts. The law and judiciary must also address novel or unusual situations where society has yet to establish clear norms. In such cases, they must resolve conflicts where existing social and ethical perspectives diverge, providing firm legal justification for their decisions[915]. Ultimately, whether the duty of care has been fulfilled will be determined by the courts based on the specific circumstances of each case[916]. In making these determinations, courts can and must consider the body of jurisprudence and scholarly literature developed on the matter[917]. In novel scenarios, particularly with emerging technologies like AI, established norms may be inadequate. Courts must balance ethical principles with technological advancements, while AI's rapid evolution and risks demand heightened due diligence (including risk analysis) from manufacturers. In cases where written legal norms do not provide clear guidelines, judges should attempt to determine whether due care was neglected by balancing the interests of individual freedom with the requirement of avoiding harm, often relying on unwritten societal rules; professional customs and common practice; and general experience-based norms to supplement legal obligations[918].

c. Under Which Perspective Should the Standard of Care Established?

In the context of negligent liability, another important issue is determining in relation to whom the duty of care should be assessed as well as identifying the legal basis of the duty of care. Indeed, individuals differ in their professions, expertise, risk perception and capacity to mitigate risks. Particularly given the unpredictable behaviour of AI-driven autonomous systems, determining the perspective from which the duty of care of the persons behind the machine is assessed, as well as whether they can legally be expected to foresee and prevent potential risks, are essential considerations. Another key consideration is whether special skills and knowledge should be taken into account. For instance, should developers at *OpenAI*

[914] DUTTGE, StGB § 15 MüKo, 2024, Rn. 33.

[915] SCHAFFSTEIN, Soziale Adäquanz, 1960, p. 394.

[916] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 156.

[917] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1188 Rn. 14; HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 556 f.

[918] HEINRICH, Strafrecht AT, 2022, p. 451 Rn. 1032; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 533 Rn. 18.

be expected to utilise knowledge possessed by only a few team members (knowledge that probably no one else in the world possesses) to reduce the likelihood of harmful outcomes produced by generative AI? If they fail to do so, should they be held liable? Addressing these questions is essential to properly establish the scope and standard of the duty of care in such contexts.

Whether negligence should be evaluated by a general or individualised standard of care has been an important point of discussion[919]. A purely objective standard imposes an unrealistic burden on the individual, while a purely subjective standard may unfairly disadvantage the affected parties by basing legal consequences solely on the individual's personal perception of danger[920]. In this context, the two-stage analysis of negligence, the individualisation theory and other perspectives offer distinct frameworks for the evaluation, each emphasising different aspects of the discussion. Nonetheless, as previously noted, they converge on broadly similar conclusions, differing only in nuanced ways, although opposing views do exist[921].

Modern mass transportation and the rise of technical risks gave rise to the need for objectifying breaches of due care, as inherently dangerous activities required precise standards to distinguish permissible risks from those deemed impermissible[922]. In this regard, the two-stage analysis of negligence begins with an objective perspective: assessing whether the risk could have been *ex ante* recognised and avoided by a hypothetical reasonable, conscientious and prudent person with the same social role as the perpetrator, using specific legal norms to define the required standard of care where applicable. This approach enables generalisation, independent of individual circumstances. In the second stage, the focus shifts to a subjective assessment under guilt, evaluating whether the specific perpetrator was personally able to recognise and avoid the risk. The individual ability to act with due care is affirmed if the offender, based on their intelligence and education (particularly their accessible knowledge of causal laws); skills; abilities; life experience and social status, was capable of recognising

[919] STRATENWERTH, Zur Individualisierung, 1985, p. 285.

[920] SCHÖMIG, Gefahren und Risiken, 2023, p. 158 f.

[921] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1201 f. Rn. 56.

[922] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 180 f. Rn. 39.

the potential consequences of their actions and could have avoided them through careful behaviour[923].

The objective evaluation under the wrongdoing (*Unrecht*) requires that the assessment should consider whether a person in the offender's position, within the relevant community, would possess the requisite knowledge and skills to manage the specific risk in question. This determination must be made based on the specific risk of the activity, thus distinguishing that group from the general public[924]. For instance, a professional is expected to possess the attributes and expertise appropriate to their field[925]. Nevertheless, application of the objective duty of care in criminal law should not dissuade individuals from exercising great caution in situations where they are capable of so doing. Similarly, it should not hinder them from exceeding the average standard or from pursuing the development of their skills and expertise[926].

Particularly in the absence of specific regulations, the importance of conducting the assessment based on a hypothetical standard figure becomes evident[927]. However, one perspective criticises this approach, asserting that it poses significant challenges in defining the appropriate reference group. Additionally, it is argued that the approach fails to offer clear guidance on the specific duties of care and a "prudent and conscientious person" would rely on an overly abstract and vague standard[928]. Another opinion criticises

[923] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 138; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 180 ff., 190 Rn. 39, 43 f., 79; STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 308 Rn. 12; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1144; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 318 Rn. 819 f.; CORNELIUS, Künstliche Intelligenz, 2020, p. 59; EISELE, §12 Die Fahrlässigkeit, 2016, p. 306 Rn. 39 f.; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 222 Rn. 16; FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 165 f., 169 Rn. 15, 24; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 266 f. Rn. 38 f.; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 532 Rn. 15.
Such context may differ, for instance, between a general practitioner and a specialist. JÄGER, Strafrecht, 2021, p. 446 Rn. 561.

[924] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 138; EISELE, §12 Die Fahrlässigkeit, 2016, p. 306 Rn. 39 f.

[925] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 182 Rn. 48.

[926] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 247 f.

[927] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1114.

[928] SCHÜNEMANN, Moderne Tendenzen, 1975, p. 575.

the two-stage analysis of negligence on the grounds that it relies on the hypothetical evaluation of a fictitious individual from the perpetrator's circle. According to this critique, each case actually involves the judgment of two individuals: one hypothetical and one real. While the foreseeability of the harm is assessed through this hypothetical person, the focus shifts to an abstract construct rather than the concrete circumstances of the case. This approach arguably disregards the specific characteristics of the actual perpetrator involved in the incident. The critique emphasises that what truly matters is whether the actual offender possessed the requisite attributes. It also highlights potential difficulties, particularly in rare cases, where the offender's unique knowledge and expertise might come into question. For example, while it may be feasible to establish a standard model for ordinary positions, defining a standard for amateurs or those in a training position poses significant challenges[929].

As noted earlier, the application of various criteria across different doctrines generally leads the two-stage analysis and other approaches to produce similar results[930]. Accordingly, the assessment of duty of care is based on *ex ante* consideration of the danger based on all relevant circumstances of each specific case. The assessment considers how a conscientious and reasonable individual within the perpetrator's social or professional sphere, possessing the perpetrator's special knowledge and skills, which could set a higher standard of care, would have acted in the specific circumstances[931]. Objective foreseeability is also a part of setting the objective duty of care. The perpetrator can only be accused of negligence if the outcome and the causal sequence were objectively foreseeable for such an individual[932], along with any additional causal knowledge they may reasonably be expected to

[929] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 168 ff. Rn. 23-27.

[930] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 181 f. Rn. 43 f.

[931] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 310 Rn. 63; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 262 f. Rn. 22 f.; VALERIUS, Strafrechtliche Grenzen, 2022, p. 124.
The assessment of whether an objective duty of care is knowable and achievable necessitates a personalised evaluation. Specifically, the standard is based on a hypothetical third person assumed to be of the same age, intelligence, cultural background, and experience as the perpetrator, placed in similar circumstances. This constitutes the subjective duty of care. See: MERAKLI, Ceza Hukukunda Kusur, 2017, p. 195.

[932] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 263 f. Rn. 27 f.

possess[933]. Case law further involves comparing the perpetrator's actual conduct to the standard of behaviour a diligent and prudent person within the same social or professional context would have demonstrated in the particular factual situation leading to the harmful outcome[934].

The evaluation of guilt for manufacturers developing and producing AI-driven autonomous systems may hold less significance, as these companies and their employees are presumed to possess sufficient expertise to create such technology. For them, the primary focus will likely revolve around the objective assessment. If an AI-driven autonomous system causes a crime, the inquiry focuses on how a careful programmer would have acted in similar circumstances[935]. However, this assessment is especially complex in novel fields such as AI. Nevertheless, in cases like the *Darknet Shopper,* a software that was developed by two amateurs, where it "accidentally" purchased illegal drugs from a darknet marketplace[936]; a subjective evaluation becomes more critical. Furthermore, the duty of care of organisations engaged in the development of AI encompasses implementing training

[933] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 587; ZIESCHANG, Strafrecht AT, 2023, p. 122 Rn. 433.
Individual foreseeability is a fundamental component of negligence-related wrongdoing, not merely of culpability. Therefore, the determination of wrongdoing hinges on the individual abilities of the perpetrator to foresee and avoid their actions in light of their statutory consequences. See: JAKOBS, 9. Abschnitt - Strafrecht AT, 1991, p. 323 Rn. 13.
In addition to the debates surrounding the two-stage analysis of negligence, the discussion about whether foreseeability and avoidability assessment in wrongdoing should be made subjectively or objectively is also crucial. The prevailing opinion advocates for an objective standard, thereby prioritising the protection of legal interests. Conversely, the minority opinion argues that these elements should be evaluated exclusively from a subjective perspective, as relying solely on objective criteria could potentially lead to a form of strict liability. For the discussions, see: GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 579 Rn. 133 ff.
Some authors who associate negligence with objective imputation also emphasise the need for subjective recognisability or individual predictability and avoidability of the disapproved risk creation. However, it is argued that such an approach is problematic, as it risks adopting a generalised assessment that disregards the specific circumstances of the case and promotes an overly standardised legal framework. For the discussion, see: DUTTGE, StGB § 15 MüKo, 2024, Rn. 106.

[934] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1159, Rn. 213.

[935] CORNELIUS, Künstliche Intelligenz, 2020, p. 59.

[936] POWER MIKE, "What happens when a software bot goes on a darknet shopping spree?", 05.12.2014, https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper. (accessed on 01.08.2025).

programmes and seminars for their developers, programmers and other relevant personnel, regarding the awareness of such potential risks, challenges, harms and legal liabilities that AI systems may pose in real-world applications.

One of the key points of debate in determining a breach of duty of care is whether the perpetrator's special knowledge and skills, as well as their general incompetence, should be taken into account[937]. The prevailing opinion asserts that, in determining negligence, such factors should be considered and individuals with greater skills and knowledge should be held to higher standards of care. The opposing view argues that care requirements should not be overstretched, particularly when risky actions serve significant social interests, and professionals; such as doctors, should not face criminal liability for adverse outcomes if they acted appropriately, unless they exhibited a gross disregard for established evaluation criteria[938]. Furthermore, it has been argued that it could lead to a double standard, and an overly subjective negligence benchmark that might result in legal complexities. Additionally imposing higher standards could deter individuals from pursuing advanced skills or knowledge, as this would indirectly enforce additional obligations on them[939]. This issue could deter companies from conducting more comprehensive risk analyses or investigating emerging risks associated with their technologies. To address this concern, it would be reasonable for the legislature to explicitly impose such obligations on these companies, thereby ensuring a proactive approach to identifying and mitigating potential risks.

The question of whether it is truly reasonable to expect individuals with remarkable capabilities to consistently demonstrate their abilities in all situations is an essential one. For instance, can a rally driver be expected

[937] Certain human abilities are significant; however, differing opinions adopt varying approaches to how these should be considered in determining negligence. An individual's instrumental and moral capacities should be assessed within the context of their personal abilities and must not be conflated with the general duty of care. See: STRATENWERTH, Zur Individualisierung, 1985, pp. 286-287.
The view that special knowledge and skills should also be considered in assessing the objective breach of the duty of care seeks to refine the evaluation of actions without contradicting the objective benchmarks typically applied to behaviour. See: KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 223 Rn. 23.

[938] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1119.

[939] SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 263.
For the evaluation, see: VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1138, Rn. 159 ff.

to drive with the same skill and precision in regular traffic as they would during a race[940]? Moreover, in a rapidly evolving field where no comparable individuals can serve as a model, using a master with unique expertise in a specific technique as the benchmark for the general standard would inevitably lead to others being deemed negligent in all cases. Therefore, maintaining consistent individualisation in the assessment of criminally relevant negligent misconduct is essential to ensure fairness and avoid unjust outcomes[941]. Negligent undertaking for overreaching capacity will be discussed further below.

According to the prevailing opinion, expecting individuals with certain technical knowledge, experience, or intelligence not to foresee and avoid the consequences of their actions would effectively create a privileged class under criminal law[942]. The average knowledge of a prudent and perceptive person pertains only to the minimum level of care and objective foreseeability. Therefore, the prevailing opinion holds that special abilities should also be considered, which is reasonable given the impracticality of distinguishing between average and exceptional abilities, as individuals inherently possess varying levels of skill[943].

[940] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 172 f. Rn. 31 ff.

[941] *Ibid.*
For example, in the case where it is investigated whether a mother who fed her child an overly salty pudding, mistaking it for sugar, could have foreseen the fatal outcome, objective foreseeability is determined not according to a doctor specialised in health; but according to an average mother in her social environment. For the example, see: HEINRICH, Strafrecht AT, 2022, p. 444 Rn. 1014.

[942] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 235.

[943] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 138; EISELE, §12 Die Fahrlässigkeit, 2016, p. 306 Rn. 39 f.; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 533 Rn. 20 f.; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 320 Rn. 824; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 191 f. Rn. 84; GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 560 Rn. 48; STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 309 Rn. 14; VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1138, Rn. 159 ff.
For the evaluation of individualisation upwards being possible if the perpetrator has special knowledge and skills, see: ZIESCHANG, Strafrecht AT, 2023, p. 122 Rn. 432.
Neither a wholly subjective nor a purely objective approach is adequate. Below-average abilities cannot exempt an individual from liability and above-average abilities must be utilised. Accordingly, the standard should be “generalised downwards and individualised upwards”. See: ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1201 f. Rn. 57.

Failing to take into account the perpetrator's specialised knowledge or skills can lead to problematic outcomes. For instance, if a doctor, through their specialised knowledge, recognises that a patient has an allergy not typically accounted for in standard medical procedures, adhering strictly to the medical *lex artis* could result in the patient's death. Therefore, the prevailing opinion asserts that the doctor is obligated to utilise their specialised knowledge in such cases[944]. Similarly, if a truck driver is aware that the cyclist ahead is intoxicated, merely maintaining the standard safety distance while overtaking would not be considered adequate[945]. Indeed, those with specialised skills, such as trained lifeguards, should be held to a higher standard, as their expertise is expected even outside their professional role[946]. A postman who becomes aware that a package contains a bomb cannot be said to fulfil their duty of care merely by "doing their job" and proceeding to delivery because criminal law addresses the individuals as law-abiding citizens[947]. Building on this example, if a programmer employed by a company happens to discover that the company's AI system (such as an LLM) processes confidential state secrets and discloses them when demanded by regular users, it cannot reasonably be expected of the programmer to remain silent and simply continue "doing their job". The same principle applies when the issue in question can only be resolved through a patch developed by the programmer themselves or their team.

d. Negligent Undertaking

The prevailing opinion supports the consideration of an individual's special knowledge and skills in determining the scope of the duty of care, as previ-

A similar approach in Turkish legal literature advocates for a modern two-stage analysis of duty of care by incorporating the offender's specialised knowledge and experience into the assessment of liability when such skills are not utilised. This model adopts a generalising approach for minimum standards while employing an individualising approach for maximum standards. As a result, it provides a tailored framework that adjusts to individuals exceeding the average level of competence. See: DEMIREL, Taksir, 2024, p. 774.

[944] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 300 Rn. 28.

[945] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 191 f. Rn. 84.

[946] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1138, Rn. 159 ff.

[947] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 410.

ously elaborated. However, it is equally important to examine the impact of below-average abilities on the offender's liability. Fundamentally, in cases of negligence, no one can be expected to exercise a level of foresight and due care beyond their capabilities. On the other hand, events can only be controlled if the individual has the ability to mitigate the risks through appropriate measures or by refraining from the risky action[948]. Therefore, this line of reasoning could lead to the conclusion that individuals lacking sufficient capacity would not bear responsibility when undertaking certain tasks, which raises critical questions on the limits of liability.

According to two-stage evaluation of negligence, the concept of subjective breach of the duty of care in criminal law assesses whether an offender can be personally blamed for their negligent behaviour. Unlike civil law, which applies an objective standard, criminal law takes into account an individual's personal attributes and abilities in the specific context under guilt. An offender is deemed guilty only if they were personally capable of adhering to the objective standard of care. If the offender lacked the requisite knowledge or skills, they would not satisfy the criteria of guilt; even though their behaviour constitutes an objective breach of the duty of care[949]. Therefore, they may not be held liable. However, there could still be grounds for negligent liability due to exceeding their capacity[950].

In such cases where an individual undertakes a dangerous activity despite lacking sufficient competence and being unable to keep the risks within permissible limits, the accusation of negligence is justified by the very fact that they chose to engage in the activity[951]. In such cases, the negligent liability arising from being, in principle, already prohibited from undertaking that activity is referred to as negligent undertaking (*Übernahmeverschulden*[952] or *Übernahmefahrlässigkeit*[953]).

Individuals should not undertake a task unless they possess the necessary knowledge and skills[954]. For example, driving at high speeds on the highway may be appropriate for an experienced driver but not for individuals

[948] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 309 Rn. 16 ff.

[949] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1201 f. Rn. 58.

[950] EISELE, §12 Die Fahrlässigkeit, 2016, p. 315 Rn. 66.

[951] HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 323 Rn. 834.

[952] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 311 Rn. 22; JÄGER, Strafrecht, 2021, p. 448 Rn. 561; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 224 Rn. 26.

[953] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 176 Rn. 40.

[954] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1117.

who may face limitations due to age-related factors[955]. In such a scenario, if an accident occurs, the perpetrator cannot evade liability due to having below-average abilities, as preventing harm remains a fundamental necessity[956].

Although self-driving vehicles aim to facilitate transportation for individuals with mobility challenges, it is essential, especially in the current era of semi-autonomous driving, to familiarise oneself with the system's requirements. Because lacking familiarity with the system and acting in ignorance by deploying and operating it, may constitute misconduct and faulty behaviour[957]. Hence, when the driving assistance system issues a warning, the driver must take control of the vehicle. If an individual, due to limitations or unfamiliarity with the system, fails to assume control and an accident occurs, they may bear liability for negligence. The basis of such negligent liability stems, in the first instance, from their decision to engage in the activity despite these limitations. Therefore, additional training could be incorporated within the scope of a driving licence to enable the use of these systems.

In my view, the most significant implication of a negligent undertaking would be a *de facto* prohibition on individuals who lack sufficient competence from engaging in the development of complex and higher risk AI systems. While this is unlikely to pose an issue for large corporations and where AI systems are developed as products; it becomes highly relevant in cases like the *Darknet Shopper*[958]. If an individual exceeds their capacity by creating an AI-driven system that is subsequently involved in criminal offences, persons behind the machine cannot evade liability by claiming their incapacity and the absence of guilt.

[955] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 308 Rn. 13.

[956] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 560 Rn. 48.

[957] VOGT, Fahrerassistenzsysteme, 2003, p. 157.

[958] It can nevertheless be argued that this instance cannot be assessed under negligent undertaking, due to the general inexperience at the time that it occurred. POWER MIKE, "What happens when a software bot goes on a darknet shopping spree?", 05.12.2014, https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper. (accessed on 01.08.2025).

e. Insights from Turkish Law on Negligence and the Scope of the Duty of Care

Negligence, while interpreted through legal doctrine and judicial practice in countries such as Germany, is explicitly defined in the criminal codes of certain jurisdictions, including Turkey[959]. Article 22(2) of the Turkish Penal Code (TPC) defines negligence as the *realisation of an act without foreseeing the consequence specified in the legal definition of the offence due to violation of the duty of attention and care*[960].

Based on the expression "realisation of an act" in this provision, it is asserted that negligence is regulated as a type of wrongdoing (*Unrecht*), which pertains to the elements of an offence (*Tatbestand*). The breach of the duty of care and foreseeability are explicitly provided for in the law. However, considering several provisions on the matter and the explanatory memorandum of the relevant provision, there are indications that negligence is structured according to a two-stage evaluation[961] or is used interchangeably with culpability in Turkish law[962]. One perspective asserts that the two-stage analysis of negligence is the prevailing approach in Turkish criminal law[963]; yet it cannot be deemed accurate considering current legal literature[964]. Case-law and the Court of Cassation has not contributed to the theoretical debate regarding the nature of negligence in Turkish law, either[965].

[959] KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, p. 183.

[960] The translation was made by the author. Although the Venice Commission has adopted the term "recklessness" to refer to negligence in English translation, this usage is inaccurate. In English legal terminology, "recklessness" aligns more closely with the German concept of *Leichtfertigkeit*, which denotes a higher degree of disregard than (conscious or unconscious) negligence. See: Council of Europe, European Commission for Democracy through Law (Venice Commission), Penal Code of Turkey, Opinion No. 831/2015, CDL-REF(2016)011, 15 February 2016, https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2016)011-e. (accessed on 01.08.2025).
For the relationship between intention, recklessness, and negligence with *mens rea* in common law systems, see: MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 57; HORDER, Ashworth's Principles of Criminal Law, 2019, p. 175.

[961] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 472.

[962] For a detailed evaluation, see: MERAKLI, Ceza Hukukunda Kusur, 2017, p. 344 ff.

[963] DEMIREL, Taksir, 2024, p. 110, 113 f.

[964] For the critique of the two-stage analysis of negligence and that it should be confined solely to the domain of wrongdoing (*Unrecht*), rather than extending into other areas, see: MERAKLI, Ceza Hukukunda Kusur, 2017, p. 351.

[965] *Ibid*, p. 350.

Unlike German law, in Turkish law, negligence is considered under the subjective element alongside intent. Nonetheless, some scholars argue that it should be examined separately, given its exceptional nature, rather than being subsumed under the subjective element[966].

There are diverse viewpoints on explaining the underlying nature of negligence. One view supports the theory of foreseeability and preventability as a coherent explanation. Accordingly, negligence is characterised by the offender's failure to foresee harmful or dangerous outcomes affecting societal order, despite possessing the capacity to do so, or by their failure to prevent such outcomes even when foreseen[967]. An alternative opinion posits that the essence of negligence lies in a breach of due care that is foreseeable in nature[968]. Another perspective contends that explaining the essence of negligence through the foreseeability theory is insufficient; mainly because it creates a contradiction in cases where an individual complies with codified behavioural rules and foresees the harmful outcome, yet they would not be held liable for negligence despite such foresight. Rather, the essence of negligence should be understood as the condemnation arising from the unintended commission of an act that could have been avoided by adhering to mandatory behavioural rules, but which occurred due to a violation of them[969].

The negligent act defined by law occurs because the required duty of care is not exercised, resulting from a failure to foresee the outcome. However, the act must have been avoidable through due care, provided that the possibility of foreseeing the outcome existed[970]. There are differing opinions regarding the position of the duty of care and foreseeability, as well as on whether these concepts should be assessed subjectively or objectively[971]. According to one view, the duty of care is objective in nature, while foreseeability is subjective. Initially, the violation of the duty of care is identified, and then the foreseeability of the outcome is assessed subjectively from the

[966] ZAFER, Ceza Hukuku, 2021, p. 343.

[967] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 471. For the explanations regarding foreseeability, see: ZAFER, Ceza Hukuku, 2021, p. 342.

[968] DEMIREL, Taksir, 2024, p. 115.

[969] TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 231 ff.

[970] ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 269.

[971] According to one view, foreseeability is examined under the concept of objective imputation in Turkish law, which is the prevailing opinion. Yet, despite a widespread acceptance, objective imputation cannot be regarded as the prevailing concept in contemporary Turkish legal literature. For the view, see: DEMIREL, Taksir, 2024, p. 378.

offender's perspective. In this context, the determination of foreseeability is based on the individual offender[972].

In determining foreseeability, one view suggests an objective standard to be applied in duty of care, whereby the assessment is based on a hypothetical person from the offender's social environment, without taking the offender's personal characteristics into account[973]. Another opinion argues that, as a rule, the standard should be that of an ordinarily prudent person. Yet, if the offender is capable of a higher due care, the determination should be made according to the offender's specific skills and knowledge[974]. An alternative view posits that the offender's personal and socio-cultural characteristics, profession, and cultural background should also be taken into account. The standard is neither that of a reasonably intelligent third party nor solely that of the offender; rather, it is a person embodying all the characteristics of the offender[975]. Another view argues that relying solely on an objective standard may lead to a strict liability regime; therefore, a mixed standard should be adopted[976].

According to the Turkish Court of Cassation, foreseeability can be explained as the possibility of an offender with specific characteristics predicting the harmful consequences of their actions. If foreseeability is impossible, the situation will instead be classified as an accident or coincidence[977]. The Court generally addresses such issues of accident and coincidence within the scope of causality, often ruling that no causal nexus exists in such cases[978]. However, it should be noted that the legal nature of accident and coincidence is a subject of debate[979]. According to the traditional view, a causal nexus exists in such cases; but the outcome was simply unforesee-

[972] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 475.

[973] ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 270.

[974] It has further been argued that the Turkish Penal Code has adopted the objective approach, although this could be contested. For the evaluation of both views, see: KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, pp. 204-205.

[975] HAKERI, Ceza Hukuku, 2022, p. 240.

[976] ÖZEN, Öğreti ve Uygulama, 2023, p. 518.
For an evaluation from the perspective of Anglo-American law, see: HALLEVY, Liability for Crimes Involving AI, 2015, p. 125 f., 134 f.

[977] Turkish Court of Cassation, General Criminal Assembly, "E. 2014/67", "K. 2016/45", 09.02.2016.

[978] HAKERI, Ceza Hukuku, 2022, p. 203 f.

[979] For the view that in such cases the outcome cannot be objectively imputed to the offender because it did not result from a breach of due care, see: KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, p. 212.

able, even under the most advanced scientific knowledge and experience[980]. This issue is significant in terms of the scope and boundaries of foreseeability, as discussed below.

In conclusion, it can be observed that the debates in Turkish criminal law literature, mainly over the past two decades, have been significantly influenced by German legal literature[981], particularly following the new Turkish Penal Code entering into force in 2005. While not entirely parallel, the discussions and practical outcomes on foreseeability and the scope of the duty of care in negligence exhibit huge similarities with the German law examined in detail above. Consequently, the *ex ante* issues discussed throughout the study in relation to crimes involving AI-driven autonomous systems remain applicable to Turkish law to the extent that their nature aligns with its legal framework.

4. The Scope and Boundaries of Duty of Care for the Person Behind the Machine

The legal nature, basis and criteria (subjective/objective) for determining liability based on negligence have been evaluated above. The primary purpose of this evaluation is to delineate the scope and boundaries of an individual's duty of care in a specific case. Indeed, with respect to AI-driven autonomous systems, the diminishing role of human control and the *ex ante* issues, primarily due to their unforeseeable nature, necessitate the establishment of clear legal parameters for determining the liability of the person behind the machine. Without such legal clarity, every harmful outcome involving these systems risks resulting in either unjustified liability or impunity.

For instance, in the objective analysis of negligence for criminal offences, such as negligent homicide that may arise in the context of self-driving vehicles, the behavioural norm regulated under Section 222 of the StGB cannot be interpreted as simply: "do not cause the death of another!" Such an imperative would be impractical to follow, given the boundless scope of the condition theory. Instead, the appropriate norm in this context should

[980] For the discussion, see: TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 249. See also: ZAFER, Ceza Hukuku, 2021, p. 463.

[981] TELLENBACH, Einführung in das türkische Strafrecht, 2003, p. 9, 2 fn.10; HEPER, Ceza Hukuku, 2019, p. 3255.

be understood as: "exercise the necessary care in the specific situation to avoid causing the death of others!"[982].

The duty of care entails considerations such as foreseeability, proactive prevention, reasonable behaviour, awareness, compliance with established standards, and avoidance of omissions when necessary. For an action to be considered a violation of the duty of care, the harmful outcome must have been both foreseeable and avoidable. An event or outcome that was neither foreseeable nor avoidable cannot lead to negligent liability[983]. The level of duty of care, as well as its connection to foreseeability and avoidability, increases in proportion to the level of risk[984].

a. The Boundaries of Foreseeability

(1) Recognising the Unforeseeable

In the context of crimes involving AI-driven autonomous systems, determining foreseeability of the outcomes is crucial for assessing whether the persons behind the machine could have avoided or prevented harm and what measures they could have taken. This analysis is essential in establishing whether there has been a violation of the duty of care. However, as detailed above[985], the autonomous nature of AI-driven systems, combined with their "self-learning" capability and adaptability, makes the foreseeability, or more broadly, the recognisability of the outcomes particularly challenging.

Within the context of this study, it is more appropriate to address not only foreseeability of the harmful outcomes, but also recognisability of the risks. Because a law-abiding individual is expected not only to avoid actions they fully foresee as dangerous; but also to identify potential risks associated with their behaviour[986]. Therefore, the duty of care should encompass not only the foresight of potential outcomes but also the responsibility to

[982] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1114.

[983] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 177 Rn. 43.

[984] HILGENDORF, Gefahr und Risiko, 2020, p. 13.
For the approach suggested in this study, see: Chapter 4, Section C(5)(b)(1)(a)(iii): "Calibrating the Duty of Care Through Risk Levels and Public Tolerance".

[985] See: Chapter 1, Section E(1): "Ex Ante: Autonomy and Diminishing Human Control".

[986] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 294 Rn. 8.

recognise risks as it involves an active commitment to conduct research to identify potential hazards. Accordingly, manufacturers must undertake careful research and empirical studies to clarify what types of malfunction and misconduct may occur[987]. For instance, as part of the required product monitoring, it is particularly important for manufacturers of self-learning systems to identify and eliminate previously unknown product risks[988].

The inherent characteristics of AI-driven systems; such as autonomy, self-learning capabilities, and adaptivity make it exceedingly difficult to predict their outcomes with precision. The self-learning feature complicates the identification of cause-effect patterns, thereby hindering the ability of operators to foresee potential risks[989]. Similarly, the adaptive nature of these systems intensifies this unpredictability by enabling them to alter their behaviour in response to changing environments or data inputs (particularly from third parties)[990]. Furthermore, the complexity of developing such autonomous systems may leave designers, developers and deployers without the necessary knowledge or capacity to anticipate the systems' conduct[991]. This unpredictability, in conjunction with their nature pushing the boundaries of determinism, can lead to unexpected and unintended consequences for the persons behind the machine[992]. For instance, in the case of a self-driving vehicle, questions arise regarding whether the individual who initiates the system and occupies the driver's seat should bear liability for an accident solely due to having started the vehicle, even if they could not have foreseen the specific chain of events leading to the harm[993]. Indeed, despite exhaustive testing to mitigate such risks, certain outcomes may still remain unforeseeable. Allowing the persons behind the machine to evade liability solely on the basis of unpredictability could lead to an unacceptable lack of accountability; effectively shielding them in almost all

[987] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 560.

[988] SANDHERR, Strafrechtliche Fragen, 2019, p. 3.
It has been suggested that the liability of manufacturers is typically reduced in circumstances where objective foreseeability presents greater challenges. However, in my view, in such circumstances, the focus should shift from foreseeability to recognisability, thereby emphasising the necessity of conducting research and development to identify potential risks. See: ASARO, A Body to Kick, 2012, p. 174.
See: Chapter 3, Section C(1)(d)(6): "Criminal Product Liability".

[989] OSMANI, The Complexity of Criminal Liability, 2020, pp. 56-57.

[990] BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 16

[991] SWART, Constructing Electronic Liability, 2023, p. 600.

[992] HAAGEN, Verantwortung, 2021, p. 220; HU, Robot Criminals, 2019, p. 513, 515.

[993] GIANNINI/KWIK, Negligence Failures, 2023, p. 51.

cases. This raises the critical question of whether all types of damage caused by such systems can or should be deemed foreseeable by the law[994].

It can be argued that, due to the probability of autonomous systems to exhibit atypical and potentially harmful conduct, users operating such systems in dynamic and complex environments must accept the possibility of occasional erroneous and atypical decisions[995]. Indeed, AI-driven autonomous systems cannot be entirely controlled; yet asserting that certain conduct is unforeseeable, because it is uncontrollable, is analogous to a zoo director releasing a tiger, and then attributing a passer-by's injury to the unpredictable nature of the animal[996].

In this regard, those who deploy, utilise, or delegate tasks to such systems must remain mindful of their inherent potential risks. Although such harmful outcomes may be infrequent, they can nevertheless materialise under certain circumstances. While the issue will be further examined within the framework of the permissible risk doctrine, it can be argued that the unforeseeability of AI-driven autonomous systems' typical risks is itself recognisable. For instance, in the case of a tiger released from a zoo, the risks it may pose are broadly predictable: it might attack a few passers-by. On the other hand, it is unlikely to simultaneously bite 100 individuals, cause a plague, or transfer personal data. In other words, typical risks are generally recognisable, and the fact that such systems cannot be controlled at every stage like puppets, does not alter this fact. Introducing these systems, along with their inherent risks, constitutes the initial anchor point for examining liability.

The identification of this anchor point is significant as it serves as the starting point for evaluating criminal liability[997]. The deployment of autonomous systems gradually diminishes human control; however, in my view, this issue bears certain similarities to the principle of *Actus Libera in Causa* (ALIC). For instance, in the case of a mother who, while sleeping, accidentally smothers her baby to death, the focus of the liability assessment lies in her actions and precautions taken before falling asleep; specifically, whether she fulfilled her duty of care through conscious and controlled behaviour prior to the loss of control during sleep.

[994] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 564.

[995] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 175.

[996] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 582.

[997] For a similar view, see: ENGLÄNDER, Das selbstfahrende, 2016, p. 374; For another similar view, see: HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 168. See also: WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 173 f.

(2) Learning from Mistakes and Hindsight Bias

Another significant issue concerning AI-driven autonomous systems is the difficulty in identifying typical or potential risks. For instance, it has become clearer from past incidents that a robot vacuum cleaner could harm an individual by pulling their hair, that a bot could engage in illicit activities such as drug trafficking, or that chatbots could insult users. Indeed, it can now be argued that manufacturers' duty of care should be elevated accordingly, given the growing awareness of the potential for such incidents to occur. Thus, they must ensure that AI-driven bots are designed to avoid engaging in harmful conduct, such as insulting users. If there are deficiencies in the programming or filtering mechanisms of these generative AI, developers may be held liable; because such harmful outcomes are now recognisable as typical risks. Assigning responsibility in this manner will urge the industry to continuously monitor and refine its technological advancements. Moreover, following incidents of this nature, the standard of care is likely to be raised incrementally, setting higher benchmarks for the development and deployment of such systems.

It should be noted that these assessments are made *ex-post*. Prior to 2015, it may not have been reasonable to expect developers of robot vacuum cleaner software to anticipate and design the system to prevent incidents such as pulling human hair, as this was not as foreseeable then as it is today. In this context, particular attention must be paid to the phenomenon known as *hindsight bias*[998], especially when determining the boundaries of the duty of care[999]. These boundaries in such innovative fields will likely be gradually defined over time through case law and experience[1000]. However, the recognisability of risks should be assessed according to the *ex-ante* characteristics of each individual case; otherwise a shift from fault-based liability to strict liability may occur[1001].

[998] Hindsight bias is the tendency to overestimate the predictability of an event after knowing its outcome, leading to the belief that the event could have been anticipated more accurately than it actually could have been. See: DAHAN-KATZ, The Implications of Heuristics, 201313, p. 153.

[999] GLESS, Mein Auto, 2016, p. 238; SCHUSTER, Künstliche Intelligenz, 2020, p. 399.

[1000] See also: Chapter 4, Section C(4)(b)(4): "The Evolution of Duty of Care Through New Techniques".

[1001] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 9.

(3) Objective Foreseeability, Typical Risks and Laplace's Demon

Foreseeability is an inherently abstract and vague concept, presenting significant challenges in its determination and proof[1002]. Particularly in the context of recently emerging technologies, identifying typical risks and determining the frequency of specific outcomes is specifically challenging. Such technologies often face a range of unforeseen challenges, that could be referred to as "teething problems". However, *ex ante*, it is rarely possible to predict the course of events with complete accuracy. As society, it will take time for us to fully comprehend the cause-and-effect correlations -if any- associated with AI-driven systems. Nevertheless, greater knowledge of the relevant facts enhances the predictability of outcomes[1003] and the more foreseeable a behaviour's potential to cause harm, the more likely it is to be considered a breach of duty[1004].

Greater knowledge of the facts enables the prediction of possible outcomes with greater accuracy, akin to the capabilities attributed to *Laplace's Demon*[1005]. However, the standard for what is recognisable should neither be equated with *Laplace's Demon* -an omniscient being- nor with the most insightful person[1006]. Moreover, an omniscient position has not yet been achieved in the field of risk assessment. The existing technological infrastructure does not permit absolute knowledge of the probability and full consequences of harm arising from decisions made by AI-driven autonomous systems. Nevertheless, this limitation does not preclude the consideration of risk assessments. In this regard, one perspective suggests that

[1002] OSMANI, The Complexity of Criminal Liability, 2020, p. 67.

[1003] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 183 f. Rn. 52 ff.

[1004] HARDTUNG, StGB § 222 MüKo, 2021, Rn. 16.

[1005] Laplace's Demon is a hypothetical construct representing an entity possessing complete knowledge of all variables and natural laws, enabling it to predict every future event and reconstruct every past event with absolute certainty in a deterministic universe. See: LAPLACE Pierre-Simon, A Philosophical Essay on Probabilities, Translation: Frederick Wilson Truscott/Frederick Lincoln Emory, New York: John Wiley & Sons, 1902, https://archive.org/details/philosophicaless00lapliala/page/100/mode/2up. (accessed on 01.08.2025).

[1006] The assessment of recognisability must be conducted from an *ex ante* perspective at the time of the act itself, excluding any information that could only be obtained through the subsequent fulfilment of the duty of care. See: VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1177 f., Rn. 259.

the *ex ante* standard for evaluation should not be based on "an observer equipped with the maximum knowledge of their time"[1007].

The imposition of liability on those who develop, manufacture, and utilise AI-driven autonomous systems to foresee all potential harmful outcomes effectively amounts to the application of strict liability, and this could lead to the inability to act when using such systems[1008]. It is impractical in everyday life to carry out every minor action with meticulous consideration of its potential consequences, as this would lead to paralysis in decision-making and action. Therefore, failure to perceive a dangerous situation constitutes negligence only if the person had a reason to be attentive, particularly if their knowledge or experience could have alerted them to the possibility of such a circumstance[1009]. For instance, giving a child a toy without thoroughly considering whether it might harm them is a common occurrence in daily life. In this context, even penalising such minor forms of negligent behaviour has been subject to criticism[1010].

The key question is whether foreseeing a general and abstract possibility of harm is sufficient to establish the negligent liability of the person behind the machine, or whether it is necessary for a specific, concretised scenario within a defined causal relationship to be foreseeable. For programmers and manufacturers, all typical potential harms that AI-driven autonomous systems might cause should be, in essence, be abstractly foreseeable[1011]. In exceptional cases, adaptive and self-deciding systems may generate outcomes that could be considered surprising; nevertheless, it can be generally expected that even such outcomes can be broadly anticipated[1012]. From a legal perspective, foreseeability relates primarily to the general likelihood of harm (for instance, the possibility of a self-driving vehicle colliding with someone), while the specific details of the situation may remain unforeseeable[1013] (*e.g.* the accident occurring due to the inability to distinguish

[1007] FELDLE, Notstandsalgorithmen, 2018, p. 126 f.

[1008] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 172; GÜNSBERG, Automated Vehicles, 2022, p. 447.

[1009] FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 172 Rn. 16.

[1010] *Ibid*, p. 174 Rn. 20.

[1011] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 447 Rn. 32.

[1012] SEHER, Intelligent agents, 2016, p. 53.

[1013] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1177, Rn. 258; BECK, Intelligent Agents and Criminal Law, 2016, p. 139; BALKIN, The Path, 2015, p. 52; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 443 Rn. 17.

the white truck against the brightly lit sky[1014]). In this manner, it can be argued that when deploying systems known to be risky, even if their specific outcomes cannot be entirely predicted, such risks may still be considered reasonably recognisable. This entails the ability to foresee the broader context of an action and to predict (at least in general terms) the consequences of that action within its context[1015]. For instance, following the *Tay* incident, it was undoubtedly foreseeable and a typical risk that a social media chatbot (*Grok*), when prompted to "not shy away from making claims which are politically incorrect"[1016], could engage in defamatory or offensive speech towards users. It is not necessary for the exact content, severity, or specific targets of the insult to be pinpointed in advance.

In this regard, the identification of typical risks is crucial in determining foreseeability[1017]. Objective foreseeability is excluded in cases involving events that fall entirely outside the scope of ordinary experience, where they cannot be reasonably expected[1018]. This principle applies particularly to atypical causal processes that deviate significantly from general life experience. German courts, while generally adopting a broad interpretation of foreseeability and requiring only that the final outcome be foreseeable (without necessitating the foreseeability of intermediate steps), make an exception for situations where the chain of events is so unusual that no one could have reasonably anticipated it, even with due care[1019]. Consequently, atypical events are deemed to lie beyond the scope of foreseeability[1020]. For instance, a traffic accident involving a self-driving vehicle constitutes a typical risk and is generally foreseeable for manufacturers. However, the vehicle's software malfunctioning and subsequently hacking into a bank's information system would be considered an atypical risk; which is, in the absence of specific knowledge, objectively unforeseeable.

[1014] KLEIN Alice, "Tesla driver dies in first fatal autonomous car crash in US", 01.07.2016, https://www.newscientist.com/article/2095740-tesla-driver-dies-in-first-fatal-autonomous-car-crash-in-us/.(accessed on 01.08.2025).

[1015] KARNOW, Liability, 1996, p. 190.

[1016] CHAYKA Kyle, "How Elon Musk's Chatbot Turned Evil", 16.07.2025, https://www.newyorker.com/newsletter/the-daily/how-elon-musks-chatbot-turned-evil. (accessed on 01.08.2025).

[1017] See: Chapter 4, Section C(5)(a)(3)(d): "Does Permissible Risk Cover Atypical Risks of AI?".

[1018] HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 320 Rn. 825.

[1019] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 587.

[1020] *Ibid*, p. 586 f.; ZIESCHANG, Strafrecht AT, 2023, p. 122 Rn. 433

The outcome is objectively foreseeable if a reasonably prudent person from the perpetrator's environment would have, under the given circumstances and based on general life experience, expected the occurrence of the outcome *ex ante*[1021]. 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[1022]. Thus, even if there is a causal link between the behaviour and the result, liability cannot be imputed for an outcome that was not objectively foreseeable[1023]. Moreover, if the perpetrator possesses special knowledge, this is also taken into consideration[1024].

The judiciary in Germany determines whether the offender could have recognised the fulfilment of the offence if they had exercised the level of care expected given the circumstances and their personal knowledge and abilities. However, the limit of recognisability is practically set by generalising based on life experience and by considering the violation of special norms as an indicator of recognisability[1025].

The question of foreseeability is easy to answer in the case of conscious negligence, because the perpetrator has at least recognised the danger, even if they have violated their duty by trusting that the result will not occur[1026]. For instance, if a manufacturer foresaw the potential for harm in the production of a highly autonomous system but failed to implement preventive measures[1027], or if an individual operates under the assumption that an autopilot system will not fail and an accident occurs, liability for conscious negligence may arise[1028].

[1021] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 300 Rn. 29; HEINRICH, Strafrecht AT, 2022, p. 444 Rn. 1014; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 226 Rn. 35; CORNELIUS, Künstliche Intelligenz, 2020, p. 60; JOERDEN, Strafrechtliche Perspektiven, 2013, p. 207

[1022] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1175, Rn. 252; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 226 Rn. 35.

[1023] JOERDEN, Strafrechtliche Perspektiven, 2013, p. 207

[1024] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 183 f. Rn. 52 ff.

[1025] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1178 f., Rn. 262; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1145.

[1026] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 587; GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 556 Rn. 23 ff.

[1027] MÜSLÜM, Artificial Intelligence, 2023, p. 141-142.

[1028] KÖKEN, Yapay Zeka, 2021, p. 269.

A significant issue concerning AI-driven autonomous systems is that, even if the cause of harm can be identified *ex post*, the harm may arise from unknown or unforeseen deviations despite the person behind the machine (e.g., the manufacturer) having taken all necessary precautions. One opinion argues that, under conditions of limited foreseeability, holding manufacturers liable for negligence would amount to penalising innocent parties. Accordingly, such incidents should be classified as 'accidents'[1029]. Undoubtedly, AI-driven autonomous systems will always involve some degree of unpredictability, and completely unforeseeable circumstances pose challenges in terms of criminal liability. However, it is essential to conduct a thorough examination before concluding that certain outcomes were unforeseeable (excluding liability), particularly for those who design and manufacture such systems. Indeed, advancements in modern science and technology facilitate the foreseeability of certain risks through appropriate risk assessment. For instance, comprehensive analyses can even predict the probability and potential consequences of natural disasters such as floods or tsunamis[1030]. Should manufacturers, therefore, be held liable for every generally foreseeable situation? The answer to this question should be negative. Otherwise, it would be impossible to sustain life full of risks. A more detailed analysis of this issue will follow, particularly concerning the concept of permissible risk.

b. Compliance with the Duty of Care: The Scope and Key Obligations

The expected diligence from the perspective of persons behind the machine encompasses both an internal dimension (recognising risks) and an external dimension -mitigating or limiting those risks through appropriate precautions[1031]. For negligent liability, it is essential to demonstrate not only

[1029] MÜSLÜM, Artificial Intelligence, 2023, pp. 143-147

[1030] According to the German Federal Court of Justice (BGH), *force majeure* is an external event caused by elementary forces of nature or by the actions of third parties, which is unforeseeable according to human insight and experience, cannot be prevented or made harmless by economically acceptable means even by the utmost care reasonably to be expected in the circumstances. (Federal Court of Justice (BGH), judgment of 23.10.1952, Case No. III ZR 364/51, reported in NJW 1953, p. 184). For the information see: HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 445.

[1031] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 299 Rn. 24.

that the risky situation could be recognised, but also that it could have been avoided. For example, during lawful driving, a child suddenly running into the path of the vehicle may be considered unavoidable[1032]. In analysing negligent offences, the first step involves identifying which individuals in the chain of developer, manufacturer, producer, or user activated the risk factor and, through their conduct, causally contributed to the harmful outcome[1033]. Clearly defining the scope of the standard of care is critically important; because the preventative function of criminal law is effective only when it is apparent which behaviours must be avoided[1034].

(1) The Anatomy of Failures in AI-Driven Systems

In events with harmful outcomes involving AI-driven autonomous systems, it is of paramount importance to ascertain the specific underlying cause(s). There are various potential grounds for failures in such systems, including software and hardware deficiencies as well as user-related factors. Software problems may include defects caused by errors, malfunctions, or an incomplete dataset, as well as incorrect data, poor design, inadequate testing, or failures in maintenance and updates. Similarly, hardware issues may stem from design or manufacturing defects, or problems with system components such as sensors or cameras. The design and installation of the system must ensure that it does not permit improper use and includes safeguards to prevent unforeseen misuse, alongside adequate warnings and documentation for users[1035]. Additionally, dependence on unverified components, inaccurate or incomplete data, or erroneous user inputs can undermine system performance. User over-reliance on AI outputs without applying

[1032] FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 173 Rn. 18.

[1033] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 564; KAIAFA-GBANDI, Artificial intelligence, 2020, p. 314.

[1034] BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 133.

[1035] For instance, in a scenario where a child leaves a tour group without authorisation during a factory visit, approaches a semi-autonomous robotic mechanism and is injured; neither the manufacturer nor the operator of the machine would be held criminally liable if it can be assumed that they were not reasonably expected to foresee that children might approach the machinery, and they took necessary precautions. However, the tour guide could be held criminally liable under Section 229 of the StGB for failing to fulfil their duty of supervision, as their negligence contributed to the incident. See: HILGENDORF, Recht und autonome Maschinen, 2015, pp. 16-17.

independent judgment further impairs risks. Errors arising during the AI's training process highlight the importance of avoiding the premature release of the product to the market. In autonomous driving for instance, failures could result from missing, incorrect, or poorly processed data. Ultimately, liability may originate from defective software (*e.g.*, a flawed object recognition programming) or hardware malfunctions[1036].

The complexity of AI-driven systems highlights the critical importance of meticulous design, testing, and maintenance processes. Even an incident, such as a self-driving vehicle causing an accident due to an improper lane change, could arise from a multitude of underlying factors. Precisely identifying the specific component failure responsible for the accident is essential to establish liability. Although this process may sometimes be hindered by issues of system opacity[1037], when the specific cause can be identified, liability can be attributed to those accountable for the faulty component -such as the provider of the dataset, the manufacturer of the sensors, or the architect responsible for the flawed and unchecked ML algorithms. Hence, the scope of the duty of care for the person behind the machine can be more clearly defined in light of these potential issues, particularly due to their obligation to mitigate risks.

(2) Challenges in Defining Standards of Conduct for Emerging Technologies

In determining the duty of care, specific comprehensive behavioural norms regarding the avoidability of harmful outcomes and risk mitigation have not yet been fully established for AI-driven autonomous systems, due to the novelty of this technology[1038]. Therefore, the persons behind the machine face challenges in assessing their duty of care[1039]. In such cases, even the question of how an experienced and prudent individual would act in technical oversight, becomes ambiguous in complex fields like robotics and remains hypothetical[1040]. Besides, despite identifiable common breaches

[1036] GERSTNER, Liability Issues, 1993, p. 248 f.; ASARO, A Body to Kick, 2012, p. 173.

[1037] See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".

[1038] STAUB, Strafrechtliche Fragen, 2019, p. 397; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 154.

[1039] ZHAO, Principle of Criminal Imputation, 2024, p. 14.

[1040] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 453; BECK, Intelligent Agents and Criminal Law, 2016, p. 139.

of duty in this field, such as errors in modelling; selecting training data, evaluating safety and the concept of proper care remains highly vague[1041]. Thus, determining which industry practices should be followed and establishing clear standards becomes challenging[1042]. In this context, in addition to considering what behaviour can be expected from a reasonable person within a particular social circle; 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[1043].

In many areas, such as road traffic, there are legal rules regarding permitted or prohibited behaviour, which at least indirectly express specific disapproval of certain actions and the permitted actions' conditions[1044]. For example, in traffic, pursuant to Sections 3(1) of the StVO and 315c of the StGB, the driver is prohibited from creating risks that could lead to a loss of control over the vehicle. Moreover, the driver must consider both objective factors such as weather conditions and personal factors, including their own conditions and abilities. This represents the individualisation of due care requirements within the framework of a general norm[1045].

The abstract principle of who a prudent and conscientious person in a specific situation and social role of the person involved is[1046], is made concrete through standards of care that mandate specific behaviours for defined scenarios. For instance, the standards of care for users of self-driving vehicles are addressed in Section 1b of the StVG. According to this provision, the duties of care imposed on the driver when using "highly or fully automated systems" are limited to monitoring the system and assuming control when necessary. As a result, the level of concentration required from the driver during the automated phases of a journey is significantly reduced[1047]. In accordance with these rules, if a driver relinquishes control to the vehicle and uses the system as intended, they are entitled to rely on the assurance that it does not pose risks beyond an acceptable level for themselves or third parties. If the vehicle's hardware or software is unsuitable or defective, resulting in an accident; the manufacturer's liabili-

[1041] FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 884.

[1042] ASARO, A Body to Kick, 2012, p. 172.

[1043] BECK, Die Diffusion, 2020, pp. 46-47.

[1044] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 178 Rn. 47.

[1045] STRATENWERTH, Zur Individualisierung, 1985, p. 296.

[1046] See: Chapter 4, Section C(3)(c): "Under Which Perspective Should the Standard of Care Established?".

[1047] STEINERT, Automatisiertes Fahren, 2019, p. 5.

ty comes into question[1048]. However, it is essential to conduct a detailed assessment of whether all relevant parties have fully met their respective duties of care in such cases.

(3) The Application of the General Duty of Care

(a) Defining the General Duty of Care

As detailed in the evaluation of the legal basis for the duty of care[1049], even in the absence of explicitly defined rules for the relevant involved parties in the context of AI-driven autonomous systems, the general duty of care undoubtedly applies. The required degree of care required is dynamic; shaped by both the probability of harm and the potential severity of its consequences, yet constrained by the bounds of reasonableness. Relying on a "careful person" standard, however, carries the risk of excessive generalisation. The specific content of a duty of care can only be determined on a case-by-case basis and determining whether harm could have been avoided requires tailoring the standard to the specific context, considering all relevant circumstances in which a careful person in the offender's position would have recognised and prevented the potential outcome. Nonetheless, particularly in the context of self-learning adaptive systems, the duty of care for developers should be confined to acting within the boundaries of their expertise and professional responsibilities. Moreover, if the perpetrator possesses special knowledge, this is also taken into consideration[1050].

Determining the duty of care is crucial in the context of difficult-to-foresee or unpredictable events. For instance, if a child suddenly runs into the road from behind a parked car and is struck by a vehicle driving lawfully at a reasonable speed, the driver cannot be expected to specifically foresee this outcome and would not be held liable. However, if the child is visible and the driver sees them, liability may arise if the driver fails to exercise greater caution, as children are known to act unpredictably[1051]. Similarly, depend-

[1048] *Ibid*, p. 6.

[1049] See: Chapter 4, Section C(3)(b): "The Legal Basis of Duty of Care".

[1050] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1181, Rn. 266a; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 183 f. Rn. 52 ff.; ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 177, 180

[1051] JOERDEN, Strafrechtliche Perspektiven, 2013, p. 208.

ing on the application area of AI-driven autonomous systems -particularly if they pose greater risks or operate with greater autonomy- persons behind the machine must maintain closer supervision and be prepared to intervene immediately when necessary[1052].

In determining the duty of care, a legal prohibition designed to mitigate the dangers would play a significant role[1053]. In a risk society, even minor negligent behaviour can lead to significant consequences; therefore, adhering to expected safety standards and failing to avoid risks can result in liability[1054]. In this regard, the performance required from an individual depends on the type and extent of the risk they are allowed to create for others' legal interests. The absence of a specific regulation or standardisation for an activity, does not absolve an individual from using all available means to prevent harm when a specific danger arises. In such cases, the individual must exercise the utmost care. For instance, a rally driver is expected to use their exceptional skills to avoid hitting a pedestrian who suddenly runs into the road; they cannot argue that an average driver would have caused an accident in similar circumstances[1055].

(b) The Duty of Care Stemming from Increasing Risks

The creation or increasing of a risk inherently imposes a responsibility to prevent any harmful outcomes that may arise from that risk. By deploying or using an inherently uncontrollable AI-driven system, the person behind the machine creates an increased risk. For example, if it is discovered that a self-driving vehicle causes harm for a particular reason (even rarely), the manufacturer is obligated to address the issue and, if necessary, recall the vehicle. This obligation arises not from a prior breach of duty or unlawful conduct, but from the legitimate assumption of the increased risk[1056].

In this context, the operator of a self-driving vehicle has a duty to monitor the vehicle as a source of danger and ensure that it is in a roadworthy

[1052] *Ibid*, p. 207, 209.

[1053] However, this should not be confused with the requirement in omission crimes to have the ability to recognise and avoid criminally relevant consequences. See: STRATENWERTH, Zur Individualisierung, 1985, pp. 292-293.

[1054] SCHÖMIG, Gefahren und Risiken, 2023, p. 82.

[1055] STRATENWERTH, Zur Individualisierung, 1985, p. 300 f.
For the same view, see: THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 285.

[1056] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 585.

condition. Similarly, the driver has monitoring obligations regarding the functionality of the (semi)autonomous[1057] vehicle before starting a journey, such as checking that sensors are not covered with ice during winter[1058]. Such precautions are crucial because risk mitigation for these vehicles is most effective before the system is initiated, while interventions after activation have limited impact but still fall within the scope of the duty of care.

The establishment of sufficient trust in the safety of such systems will necessitate a length of time, during which the necessity for personal monitoring will remain[1059]. Unless a system operates fully autonomously, it remains under the partial control and supervision of the person deploying it[1060]. For example, if a parking assistance system is utilised and a child playing in the parking area is injured because one of the vehicle's sensors was dirty, this falls within the scope of due care of the driver. In such specific incidents, foreseeability and avoidability are examined[1061]. In light of the increased risk, autonomous systems should not be used as a means for individuals to evade responsibility[1062]. Delegating a task that would normally be performed by an individual and then claiming a lack of control or involvement is an inadequate defence[1063].

(c) Obligations Arising from System Failures

Another obligation that can be derived from the general duty of care is the operator's obligation to exercise greater caution when the system begins to behave unusually. Anyone with extended experience using a system is expected to recognise when it is not functioning correctly and act accordingly. To illustrate, in the case of a self-driving vehicle that typically functions properly but begins to behave abnormally, this signals a potential

[1057] The term "(semi)autonomous vehicle" refers to both semi-autonomous and fully autonomous vehicles.

[1058] VALERIUS, Sorgfaltspflichten, 2017, p. 14 f.

[1059] HILGENDORF, Moderne Technik, 2015, p. 103.

[1060] It can even be argued that delegating a task to fully autonomous systems can also be evaluated based on the conditions of such deployment and the responsibilities involved at that point.

[1061] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 803.

[1062] The impact of increasing risk on liability is examined in detail below. See: Chapter 4, Section C(5)(b)(3)(b): "Risk Enhancement through Task Delegation to AI-Driven Autonomous Systems: A Legal Analysis".

[1063] GLESS, Mein Auto, 2016, p. 243, 250.

malfunction. Taking over control only at the moment of malfunction would possibly be too late. In such a situation, the driver is required to intervene or take control immediately (as soon as they notice the abnormally); failing to do so would constitute a breach of the duty of care[1064]. It should be noted that this general duty of care is explicitly formulated in Section 1b of the StVG, but even in the absence of such regulation, it could be derived from the general principle of harm avoidance. Furthermore, to intervene effectively in dangerous situations and avoid negligent undertaking, the driver or operator of an AI-driven autonomous system must adequately familiarise themselves with its functioning. Failure to do so and behaviour contrary to the obligations outlined in the system's manual, could give rise to negligence[1065].

The decisive point of intervening would be whether the operator recognises that the technology is about to fail and that there is a need to intervene. Determining the circumstances that necessitate intervention in the operation of an AI-driven autonomous system and the assumption of control is a critical issue. Because intervening under the wrong circumstances may also result in a failure of properly performing due care[1066]. If such awareness is not possible, the operator is entitled to rely on the technology, and the manufacturer's liability may come into question[1067].

Negligent omission may be established in certain criminal offences, such as negligent homicide or bodily harm, involving AI-driven autonomous systems, particularly when a legally obliged person fails to act despite being required to do so. The party deemed negligent is typically held liable for failing to recognise a dangerous situation, for failing to assess the available options to prevent harm, or for choosing an ineffective response in accordance with Section 13 of the StGB. Liability arises if -according to the circumstances- it is established that the harm could have been prevented

[1064] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 803; SCHUSTER, Künstliche Intelligenz, 2020, p. 395.

[1065] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 147.

[1066] See: Chapter 4, Section C(5)(b)(3)(c): "Does the Non-Use of AI-Driven Autonomous Systems Breach the Duty of Care?" and Chapter 4, Section C(4)(d): "Control Dilemma".
Holding a driver liable both for failing to intervene and for intervening at the wrong moment violates the principle of guilt. See: THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 28.

[1067] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 288.

through proper action, provided that no external factors undermine this causality[1068].

(d) Duty to Ensure Robust System Design

In the context of AI-driven autonomous systems, different parties bear distinct duties of care. As operators increasingly lose direct control, it shifts toward the system's activation, design, and production stages. For example, in the context of self-driving vehicles, violations increasingly arise from the failure to perform maintenance, inspections, or properly taking control when necessary[1069]. Indeed, as the level of autonomy increases, determining the duty of care expected from the operator will become increasingly challenging[1070]. In highly autonomous vehicles, it is argued that the individual inside the vehicle transitions from the role of 'driver' to that of 'passenger'; with control and responsibility shifting entirely to the manufacturer. Consequently, misconduct in driving is being replaced by liability for product defects[1071]. Accordingly, passengers can only prevent accidents by choosing not to initiate the vehicle at all[1072].

A significant question that arises is whether the design of AI-driven autonomous systems to be resilient to third-party attacks falls within the scope of manufacturers' duty of care[1073]. Since such vulnerabilities can expose both users and third parties to significant risks and often result in criminal offences; these systems must be designed with a certain level of robustness against such attacks. For instance, Section 1f(3) of the StVG emphasises the importance of designing and producing systems capable of withstanding cyberattacks, thereby imposing specific obligations on manu-

[1068] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1093 f., Rn. 62; WEIGEND, § 13 Begehen durch Unterlassen in LK, 2020, p. 939, Rn. 97.

[1069] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 179.

[1070] BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 19.

[1071] HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 25; HILGENDORF, Wer haftet für Roboter? Autonome Autos. In: Legal Tribune Online (LTO), 21.07.2014; HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 24; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 286, 289; LOHMANN, Liability Issues, 2016, p. 337; SCHUSTER, Künstliche Intelligenz, 2020, p. 396

[1072] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 169; MÜSLÜM, Artificial Intelligence, 2023, p. 156.

[1073] HILGENDORF, Digitalisierung, Virtualisierung und das Recht, 2020, p. 417.

facturers in this regard[1074]. Indeed, even where a product does not itself cause harm, a failure to provide the protection it purports to offer, or which users may reasonably expect it to afford, may give rise to a breach of the duty of care. This is reflected from product liability aspect in Art. 7(2)(f) of the new PLD, which provides that "relevant product safety requirements, including safety-relevant cybersecurity requirements" shall be taken into account in the assessment of defectiveness.

No technology can be completely secure. For this reason, major technology companies like Apple use bounty programmes to mitigate security vulnerabilities and other threats[1075]. The foreseeability and preventability of such threats place an obligation on the producing companies to take appropriate preventive measures. This is particularly significant in the case of cyberattacks that could be avoided with better programming, as the responsibility of manufacturers in such scenarios is more effectively identifiable. However, even with all countermeasures, successful attacks may still occur, as no technology can ever be 100% secure. Even neural implants can be hacked[1076]. Moreover, as these systems operate while connected to a network, the risks are amplified to a massive scale[1077]. In this context, the concept of permissible risk defines the boundaries[1078].

In addition to the vulnerabilities inherent in traditional computing systems, AI (-driven) systems face a wide range of unique threats due to their distinctive characteristics. Attacks aimed at exploiting, deceiving, or manipulating such systems are often evaluated under the concept of adversarial machine learning attacks. There are numerous types of adversarial ML attacks. Three main categories are: 1- fooling, which involves manipulating a trained classifier or detector during the inference phase to incorrectly classify or identify an input; 2- poisoning, where the training phase is distorted to induce specific errors during inference; 3- model inversion,

[1074] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 451.
EVAS Tatjana, European Parliamentary Research Service, Impact Assessment and European Added Value Directorate, European Added Value Unit, A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles: European Added Value Assessment, 2018, p. 26.

[1075] https://security.apple.com/bounty/.(accessed on 01.08.2025).

[1076] LIN, Why Ethics Matters, 2016, p. 79.

[1077] CHANNON/MARSON, The Liability for Cybersecurity, 2021, p. 7

[1078] HILGENDORF, Moderne Technik, 2015, p. 105.

which entails extracting data, sometimes sensitive or protected, from a trained model[1079].

Through these attacks, various outcomes can be achieved, such as causing self-driving vehicles to accelerate and crash, deceiving face recognition systems, extracting sensitive data from large language models (LLMs), and even exploiting integrated AI systems in databases through prompt injections, enabling a wide range of abuses[1080]. To combat such attacks, developers should employ, *inter alia*, techniques such as red-teaming, domain adversarial training, synthetic data generation, active learning, and regular audits to ensure robust and high-quality model performance[1081]. These measures can be considered within the scope of manufacturers' duty of care.

It is also imperative that manufacturers and developers recognise the inherent dangers of unpredictable software and implement measures to restrict its interaction with the public until it has undergone comprehensive testing in a controlled environment. Following a limited release, they must provide transparent information to customers, users, and the relevant people, not only regarding the advantages of software that evolves during use, but also the potential vulnerabilities posed by unpredictable changes in behaviour[1082]. Moreover, all tests and risk analyses serve only to mitigate risk; they cannot eliminate it entirely. Unexpected events can always occur[1083].

(e) The Protective Purpose of the Norm

To establish negligent liability, two additional considerations, *inter alia*, must be addressed: first, there must be a connection between the resulting

[1079] EVTIMOV, et al., Is Tricking a Robot Hacking, 2019, p. 900; European Union Agency for Cybersecurity, Artificial Intelligence and Cybersecurity Research: ENISA Research and Innovation Brief, 2023, p. 24.
For a study on the criminal implications of these attacks, see: KATOĞLU/ALTUNKAŞ/KIZILIRMAK, Yapay Zekâ, 2025, *passim*.

[1080] For detailed information on *adversarial ML attacks*, see: YIN, Ginver: Generative Model Inversion Attacks, 2023, p. 2123; CARLINI/WAGNER, Audio Adversarial Examples, 2018, p. 1, 6; SZEGEDY et al., Intriguing Properties, 2014, p. 4; SHARIF, et al., Accessorize to a Crime, 2016, p. 1530; SHOKRI, et al., Membership Inference Attacks, 2017, p. 3.

[1081] OpenAI (Markov et al.), A Holistic Approach, 2023, p. 15016.

[1082] WOLF/MILLER/GRODZINSKY, Why We Should Have Seen That Coming, 2017 p. 11.

[1083] HAAGEN, Verantwortung, 2021, pp. 221-222.

harm and the protective purpose of the norm that serves as the source of the duty of care. Second, the offender's breach of this duty must have created an unlawful risk, leading to the factual outcome. If this connection cannot be established (the factual outcome would have occurred even if the offender had not breached the duty of care), the principle of *in dubio pro reo* applies[1084].

An individual's failure to act in accordance with the behavioural rules prescribed under a specific duty of care, even if the outcome has occurred, does not always result in negligent liability. Outcomes that fall outside the specific protective purpose of the norm are excluded. Negligent liability arises only in relation to the outcomes the norm was specifically aimed to prevent. This connection, referred to as the protective purpose of the norm, must be applied in line with the *ratio legis* of the relevant provision. Thus, individuals cannot be held liable for extraordinary, abnormal, or purely coincidental outcomes. Mere coincidence between the conduct and the definition of the criminal offence is insufficient for liability, if the act does not fall within the protective purpose of the norm[1085].

For instance, a frequently cited example in literature illustrates this perspective: although one could argue that a driver's over speeding in town A caused the accident in town B by making them arrive at the accident site sooner, this reasoning does not align with the purpose of speed limits. A speed limit aims to prevent accidents and danger in the specific area where it applies, not to control arrival times; therefore, a driver cannot be held criminally liable for negligence[1086]. To illustrate this point further, in the event that an individual operating a motor vehicle under the influence of alcohol encounters a cyclist who makes an unavoidable and sudden left turn, resulting in a fatal accident, the driver cannot be held liable for negligence if the accident was not causally related to their intoxication[1087].

[1084] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 264 f. Rn. 30-33; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 321 Rn. 827 f.

[1085] HARDTUNG, StGB § 222 MüKo, 2021, Rn. 19; ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 365; ZAFER, Ceza Hukuku, 2021, p. 351.

[1086] KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 231 Rn. 57.

[1087] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 311 Rn. 25.

(4) The Evolution of Duty of Care Through New Techniques

When determining the scope of an individual's duty of care, new possibilities and advancements are also taken into account. For example, in medicine, a physician's therapeutic freedom is limited when a new, less risky method is available, the use of which is considered a duty of care according to current scientific standards, making the use of outdated procedures a potential basis for liability due to medical malpractice[1088]. Similarly, new methods can shape the establishment of standard of care, raising the question of whether a driver should be held liable for failing to activate a superior autonomous driving system that could have prevented an accident[1089].

To illustrate, as demonstrated in the *Aschaffenburg* incident, a driver may suffer a medical emergency during assisted driving, resulting in a complete loss of control. At that time, while the issue of the manufacturer's negligent liability was being debated, it can be argued that it could not reasonably have been expected for a lane-keeping system to incorporate a security measure that would halt the vehicle when the driver fainted. Accordingly, the manufacturer's duty of care can be considered to have been fulfilled in light of the technological standards of that period. Accordingly, it can reasonably be deduced that criminal liability would not have been incurred, given that these issues were not fully comprehended and largely unforeseeable at the time. However, the necessary measures to prevent harm in such foreseeable situations today fall within the manufacturer's duty of care, requiring the vehicle to be designed to autonomously proceed to a minimal-risk condition[1090]. Indeed, modern vehicles are equipped with technology that allows them to autonomously take control in such situations[1091]. Similarly, other past incidents such as the Darknet Shopper, robot vacuum cleaner malfunctions, and offensive chatbots contribute to shaping contemporary measures and refining the scope of the duty of care[1092].

Further illustrations on regarding the importance of adopting innovative techniques to mitigate risks can be observed in the context of self-driving vehicles. Indeed, equipping self-driving vehicles with a large number of sensors -such as LIDAR, radar, cameras, and other technologies- can

[1088] BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 124.

[1089] SANDHERR, Strafrechtliche Fragen, 2019, p. 2.

[1090] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 555.

[1091] NGUYEN, et al., Development, 2017, p. 670.

[1092] See: Chapter 4, Section C(4)(a)(2): "Learning from Mistakes and Hindsight Bias".

significantly reduce the likelihood of accidents. However, such measures may not always be economically viable and, as in the case of certain companies, may be excluded from vehicles for various reasons including economic viability and aesthetic considerations. Nevertheless, if it can be demonstrated that an accident would not have occurred had a LIDAR sensor been installed, rather than relying solely on camera, negligent liability could arise. This is because manufacturers are obligated to mitigate the risks associated with such high-risk technologies to an acceptable level. They cannot justify avoiding the implementation of risk-reducing measures, such as advanced sensors, especially in high-risk systems, on grounds of profit-maximising aims or aesthetic preferences. Therefore, releasing self-driving vehicles into traffic without equipping them with *state of the art* technologies like LIDAR, radar and others, which could make these vehicles significantly safer, may not be considered as maintaining risk within a permissible level. For instance, while self-driving vehicles that rely solely on cameras might be 90% safer than human drivers, if the addition of other sensors could raise this safety margin to 95%, such technologies must be utilised. Empirical data should form the basis for determining the extent to which these methods enhance safety.

The Wall Street Journal has recently produced a documentary highlighting significant safety concerns related to Tesla vehicles. According to the documentary, Tesla has reported over 1000 accidents to federal regulators since 2021, with hundreds of these incidents occurring while the autopilot system was active. Specifically, the documentary reveals that 44 of these accidents involved the autopilot system suddenly swerving, while 31 incidents occurred when the system failed to stop or yield for an obstacle in its path. Some of these accidents, supported by video evidence, were attributed to the inability of Tesla's software to classify obstacles captured by its cameras. For instance, the system failed to identify an overturned truck because it had not been trained to recognise such scenarios, resulting in the vehicle driving directly towards the obstacle. The documentary includes the following critical observation confirming the assessment above: "Video and data gathered from these crashes by the Wall Street Journal show that Tesla's heavy reliance on cameras for its autopilot technology, which differs from the rest of the industry, is putting the public at risk"[1093]. Indeed, Tesla's autopilot technology relies primarily on camera-based computer vision,

[1093] The Wall Street Journal, "The Hidden Autopilot Data That Reveals Why Teslas Crash", 13.12.2024, https://www.youtube.com/watch?v=mPUGh0qAqWA.

with radar serving as a backup in certain models. By contrast, other manufacturers integrate radar computer vision, and LIDAR technology in their systems, which is expensive[1094]. Tesla asserts that its autopilot system is generally much safer than human drivers and has the potential to save numerous lives. However, the claim of overall safety is insufficient; it should be emphasised that such a standard does not absolve manufacturers of responsibility. AI-driven autonomous systems, including self-driving vehicles, do not merely reduce risks; they substitute them[1095]. Indeed, there may be instances where such systems have prevented accidents that would likely have occurred due to the insufficiency of human reflexes in comparable circumstances. On the other hand, while these systems may cause fewer overall accidents, they are prone to making specific, elementary errors that humans are unlikely to make, sometimes resulting in hazardous or fatal outcomes, as demonstrated[1096]. Given these risks, employing additional sensors and designing a system to ensure their interoperability to mitigate the dangers posed by these inherently high-risk technologies falls within the duty of care. If empirical evidence supports the conclusion that relying solely on cameras for autonomous driving systems is inadequate (as the documentary suggests, with experts noting the flaws in computer vision technology and predicting its eventual obsolescence) then manufacturers must adhere to such findings. Economic or aesthetic considerations cannot justify decisions that compromise public safety[1097].

Finally, it should be stated that the required degree of care is not static and must be measured by the likelihood and severity of potential damage. However, it is not without limitations; being constrained by the permissible risk and principle of reliance. According to the permissible risk doctrine,

[1094] Without endorsing any specific company or claiming their enhanced safety, for a comparison with another company's self-driving vehicle with multiple sensors, see: https://swipefile.com/waymo-vs-tesla-sensor-suite. (accessed on 01.08.2025).

[1095] This issue will be elaborated upon below. See: Chapter 4, Section C(5)(b)(3)(a): "Substituting Existing Risks".

[1096] In fact, numerous incidents reported by users reveal that these vehicles have committed basic errors that human drivers would arguably never make. For a few illustrative examples, see: https://x.com/missjilianne/status/1869565434481221879?s=12; https://x.com/thedooberhead/status/1869502131897782451?s=12; https://x.com/factschaser/status/1916623655129305491?s=12. (accessed on 01.08.2025).

[1097] See also: OVERBERG Paul/SCOTT Emma/MATT Frank, "Inside the WSJ's Investigation of Tesla's Autopilot Crash Risks", 31.07.2024, https://www.wsj.com/business/autos/tesla-autopilot-crash-investigation-997b0129. (accessed on 01.08.2025). For a list compiling some of Tesla's such accidents, see: https://en.wikipedia.org/wiki/List_of_Tesla_Autopilot_crashes. (accessed on 01.08.2025).

the benefits of certain technical products may be so significant that some degree of damage is considered acceptable. Indeed, in reality, almost all events are at least hypothetically foreseeable, including the unexpected crash of an airplane or the sudden failure of a vehicle's brakes. Moreover, nearly all risks can be theoretically avoided by taking no action (for instance, by refraining from leaving home). Consequently, in determining whether negligence can be established, it is essential to consider whether the associated risks of harm are legally required to be avoided[1098]. For instance, if a driver has adhered to the manufacturer's instructions, fulfilled all monitoring and maintenance obligations, complied with both written and unwritten traffic rules, and driven cautiously to manage the risks inherently associated with operating a vehicle, they cannot be held liable for breaching the duty of care[1099]. Permissible risk lies at the core of this study and will be examined in detail below.

c. Human in the Loop

Artificial Intelligence-driven systems are capable of implementing decisions autonomously in certain areas, while in others, they require an approval mechanism to execute those decisions. In contexts where critical judgements are implemented, it is inherently wrong to entirely exclude human moral agents from the decision-making process[1100]. The inclusion of a "*human-in-the-loop*" is essential in AI-driven autonomous systems to ensure that human judgment and accountability remain central to decision-making processes, particularly in situations involving ethical and legal concerns. As autonomy in technology enhances, maintaining human oversight and involvement helps prevent potential detachment from the realities of the world and upholds responsibility for the conduct of these systems[1101].

[1098] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 177 f. Rn. 44 f.

[1099] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 803; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 173; STAUB, Strafrechtliche Fragen, 2019, p. 397.

[1100] ANDERSON/WAXMAN, Law and Ethics, 2013, pp. 14-18; ZUREK/KWIK/VAN ENGERS, Model of a Military Autonomous Device, 2023, p. 15.
The integration of AI with one or more human agents to form a hybrid multi-agent interaction model is widely regarded as a promising opportunity for the future in this field. See: CORNELIUS, Künstliche Intelligenz, 2020, p. 63.

[1101] HILGENDORF, Modern Technology, 2017, p. 31 f.

The concept of ***human-in-the-loop*** refers to a framework in which human involvement is indispensable to the decision-making and implementation process. In this model, the AI system provides guidance or recommendations, but human approval or action is required for implementation. Closely related is the concept of ***human-over-the-loop***, which describes a scenario where a human oversees the AI system's operations, primarily in a supervisory capacity, with the ability to intervene or modify parameters in case of unexpected outcomes or to optimise performance. By contrast, ***human-out-of-the-loop*** refers to a fully autonomous model where the AI system operates independently, making decisions without human intervention or oversight, relying solely on its programming and analytical capabilities[1102].

Ensuring human involvement in approving critical decisions provides safeguards both for maintaining the integrity of the system and for preventing harmful outcomes[1103]. However, in practice, there is a risk that, over time, reliance on automated or autonomous systems and their "recommendations" may increase, gradually shifting decision-making authority from humans to the systems; which is an issue already observed in other fields[1104].

Particularly in the field of medicine, the recommendations of AI systems, which are successful at pattern recognition, should not be followed blindly. Instead, they should be utilised merely as a supportive tool to aid decision-making. Ultimate responsibility and critical judgment should remain with human professionals. Failure to maintain critical oversight carries the risk of unquestioningly relying on opaque systems due to practical necessities in various fields, ranging from border security to preventive policing. Such reliance could lead to the widespread perpetuation of recurring biases or errors, which undermines fairness and accountability.

Finally, it can be argued that enabling the integration of humans and machines not through analogue means but via direct neural connections would introduce a new paradigm to both the concept of *human-in-the-loop* and the issue of liability. However, this topic lies beyond the scope of the present study.

[1102] Personal Data Protection Commission of Singapore, "Model AI Governance Framework (Second Edition)", 21.01.2020, https://www.pdpc.gov.sg/-/media/%20Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf, p. 30, [para. 3.14]. (accessed on 01.08.2025).

[1103] IBM Technology, "What Is a Prompt Injection Attack?", 30.05.2024, https://youtu.be/jrHRe9lSqqA?t=474. (accessed on 01.08.2025).

[1104] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 453.

d. Control Dilemma

The control dilemma refers to the expectation that the person seated in the driver's seat remains prepared to take over control of the vehicle in response to potential issues that may arise during semi-autonomous driving. Although the purpose of an autonomous system is to relieve the driver of the driving task, the obligation to monitor and control the vehicle to minimise risks causes tension[1105]. Regardless of whether the obligation to monitor and control is technically necessary, it may also be legally required under the applicable laws of a given country[1106]. Accordingly, allowing a driver to completely disengage from monitoring the vehicle while it is travelling at high speeds cannot be considered within the scope of permissible risk under current standards. This is because it creates a significant risk and, above all, contravenes established written rules, such as Section 1 of the StVO[1107].

Since AI-driven autonomous systems such as self-driving vehicles are relatively new, potential malfunctions cannot be clearly foreseen in advance. Consequently, it is reasonable to expect the intervention of a *human-in-the-loop*; namely the driver who is expected to assume control and address any issues or unforeseen events that may arise. Although this view is widely accepted, the other side of the coin reveals that, in practice, such intervention may not always be feasible due to time or situation-specific reasons

[1105] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 4; HILGENDORF, Automatisiertes Fahren und Recht, 2015, p. 67 f.

[1106] HILGENDORF, Moderne Technik, 2015, p. 102.

[1107] *Ibid.*
Furthermore, the provisions in Section 1b(1) of the StVG, which grants the driver the right to divert attention, and Section 1b(2)(2), which provides the duty to monitor, have been criticised for creating ambiguity concerning the obligations of the human driver. See: WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 77, 79.
Article 8(6) of the United Nations Convention on Road Traffic stipulates that a vehicle driver must minimise any activity unrelated to driving and, under no circumstances, use a mobile phone while the vehicle is in motion. This provision implies that the driver is still expected to be involved in the driving process. Consequently, for highly automated and fully autonomous vehicles to operate on public roads, amendments to the provision of the Convention are necessary. For the discussion, see: AKSOY RETORNAZ, Otonom Araçlar, 2021, p. 335.
For the UN Convention, see: United Nations Economic Commission for Europe (UNECE), Amendments to the Vienna Convention on Road Traffic of 1968 (Article 8, Paragraph 6), 2003, https://unece.org/DAM/trans/doc/2003/wp1/TRANS-WP1-2003-01r4e.pdf. (accessed on 01.08.2025).

that could impede the ability to override the system[1108]. Indeed, there are instances where the timeframe for intervention is so limited that such an expectation becomes practically impossible. Requiring intervention under such circumstances would constitute a violation of the principle *ultra posse nemo tenetur*; no one is obligated to do the impossible[1109].

Moreover, this obligation has been criticised on the grounds that it can shift liability from manufacturers to drivers by placing the burden of liability on individuals who are expected to always monitor their travel by keeping their hands on the steering wheel or remaining ready to take-over, even though the AI-driven system remains in control until the moment of an accident. This approach risks turning partially passive drivers into *scapegoats* while absolving manufacturers of their accountability[1110]. While human oversight is essential to address the errors of such systems, particularly during transitional periods; in my view, this issue extends beyond self-driving vehicles and encompasses all autonomous systems, posing a significant risk of scapegoating. The legal framework must approach this matter with caution to ensure liability is fairly and appropriately assigned.

It is widely criticised that requiring the driver to remain constantly attentive negates the convenience sought to be achieved with self-driving vehicles. Expecting an individual to monitor the vehicle with full attention, as if they were personally driving or controlling it, is unreasonable and undermines the very purpose of autonomous driving[1111]. Furthermore, a

[1108] LOHMANN, Erste Barriere, 2015, p. 137 f.

[1109] THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 28; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 281.

To illustrate with a recent case; the autonomous feature while performing a reverse parking manoeuvre, suddenly accelerated and collided with the vehicle behind. In such situations, even if the driver exercises due care, they have no practical opportunity to intervene. See: https://youtube.com/shorts/7_oxA0-tlE4?si=Ol5qeCrrA5TsGDs3. (accessed on 01.08.2025).

Two real-life scenarios in which the person behind the wheel was able to regain control through an instantaneous manoeuvre: https://x.com/missjilianne/status/1869565434481221879?s=12; https://x.com/thedooberhead/status/1869502131897782451?s=12.

Another example of a situation in which such intervention was almost impossible: https://x.com/factschaser/status/1916623655129305491?s=12. (accessed on 01.08.2025).

[1110] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 288.

[1111] BECK, Das Dilemma-Problem, 2017, p. 129; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 289.

Another criticism is that imposing greater duties of care does not necessarily lead to increased safety. Given that the vast majority of traffic accidents stem user cannot always anticipate how the autopilot might (erroneously) interpret and respond to a dirty traffic sign. However, if the vehicle issues a warning, the user will then become aware of such risks. The driver's duty of care should be defined as maintaining readiness to respond to warnings issued by the self-driving vehicle and to intervene immediately if a danger is perceived, provided that there is no reason to doubt that the vehicle is functioning as intended[1112]. However, even in this scenario, the system must issue the warning within a reasonable timeframe; otherwise, such a requirement would conflict with the principle of *ultra posse nemo tenetur.*

5. The Permissible Risk Doctrine

a. Conceptual Framework

(1) The Concept of "Permissible Risk"

Throughout the study, the term '*permissible risk*' has been adopted to correspond to the German legal concept of *erlaubtes Risiko*. Although this concept is not widely prevalent in English legal literature, this choice aligns with the terminology commonly used therein, rather than alternatives such as acceptable risk[1113] or similar expressions[1114].

To better understand this concept, it is essential to comprehend the dynamics of the extensive industrialisation that characterised the late 19th century. During this period, industrialisation led to a significant increase in the number of individuals working in mines and factories, where they faced severe dangers to life and limb. Remarkably, in the final quarter of the 19th century, the *Reichsgericht* adjudicated numerous cases of negligent homicide or personal injury occurring in industrial plants, largely due to

from human error, requiring constant monitoring and intervention from drivers could even have the opposite effect. See: THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 29.

[1112] GLESS, Mein Auto, 2016, pp. 235-236; KANGAL, Yapay Zeka, 2021, p. 136.

[1113] The authors have adopted the term "socially acceptable risk". See: GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 434.

[1114] Those using 'permissible risk': BOHLANDER, Principles of German Criminal Law, 2009, p. 55, 97; VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 669; LEITE, Self-Driving Cars, 2024, p. 144.
The author uses "permitted risk" rather than "permissible". See: ZHAO, Principle of Criminal Imputation, 2024, p. 77 ff.

inadequate protective equipment and insufficient technical safeguards[1115]. Furthermore, it has been stated that in 1861 the Munich Court of Appeals determined that railway operations were unlawful due to the potential dangers involved. However, by the late 19th century, it was acknowledged that some risks must be tolerated to enable the utilisation of new technologies[1116].

It can be said that the rapid industrialisation posed a dual challenge. On the one hand, it brought about significant advancements in welfare and economic opportunities, while on the other, it gave rise to serious risks that demanded careful management. This critical tension, although not explicitly termed "permissible risk" was addressed by *Carl Ludwig von Bar* as early as 1871. Accordingly, there are certain dangerous; but beneficial operations, which are indispensable as they meet certain needs in our lives. However, it can be statistically foreseen that over an extended period and through the occurrence of various events, a certain number of individuals will suffer harm and even lose their lives[1117]. Subsequently, in 1895, *Alexander Löffler* proposed that risky actions should be permitted, provided that the public interest in undertaking them outweighed the associated risks[1118]. Later, *Karl Binding* conceptualised the term in 1919, emphasising that certain behaviours that provide societal benefits inevitably involve risks; but since the only way to avoid these risks is to refrain from such behaviours, individuals should not be blamed for these risks[1119].

Due to the progress in technology and science, the understanding of danger and risk[1120] evolves. Danger, which used to be perceived as originating in nature, now finds its source in "dangerous things"[1121]. Indeed, following the Industrial Revolution, many risks previously posed by natural causes were mitigated. However, with the introduction of human-made machinery into daily life, numerous previously unknown risk factors also

[1115] PREUß, Untersuchungen zum erlaubten Risiko, 1974, p. 15 f.

[1116] SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 257.
The author of this study was unable to personally confirm this information.

[1117] von BAR Carl Ludwig, Die Lehre vom Kausalzusammenhang im Recht, besonders im Strafrecht, 1871, p. 14.

[1118] LÖFFLER, Die Schuldformen Des Strafrechts, 1895, p. 8 fn. 4.

[1119] See: BINDING, Die Normen und ihre Übertretung, 1919, p. 433 ff., 441 ff.

[1120] For a terminological explanation of the concepts danger and risk see: HILGENDORF, Gefahr und Risiko, 2020, p. 11 ff.

[1121] FISCHER, Gefährliche Sachen, 2020, p. 142.

emerged[1122]. Therefore, when referring to *permissible risk*, the term "risk" refers to human-made hazards, not the natural disasters[1123].

Requiring individuals to always investigate the potential consequences of their actions before acting is unrealistic, as it would make nearly every behaviour appear negligent and prevent practical decision making[1124]. Adhering to the required standard of care does not necessitate avoiding all behaviour that could potentially limit the prevention of harm; indeed, it is not even feasible. Instead, society relies on taking calculated risks within socially acceptable levels. Engaging in risky activities is generally not deemed a breach of due care, provided that the relevant standards of care or safety rules relevant to the particular field are observed[1125].

It is important to recognise that innovations, such as AI-driven autonomous systems often come with inherent risks. It is often the harm they initially cause that drives further improvements to that technology[1126]. Inevitably, statistically at some point, injuries will occur. In this context, criminal liability can only be avoided if such systems are never manufactured in the first place[1127]. Although new technologies aim to mitigate already acknowledged risks, absolute safety in all situations cannot be guaranteed. No manufacturer or regulatory body can anticipate every possible interaction between an adaptive system and human actors across all conceivable scenarios[1128]. Therefore, certain actions, despite their risky nature are permissible if appropriate safety measures and standards of care are observed. These actions, although inherently dangerous, do not lead to criminal liability as long as the necessary precautions are taken[1129].

One might question whether the term *permissible risk* refers solely to the authorisation of a risky activity, and thereby does not cover the harm materialising from that risk. For instance, the operation of self-driving vehicles constitutes a highly risky activity, and legal systems typically restrict or prohibit such activities. In this regard, when assessed within the framework of permissible risk, it is entirely reasonable to argue that while the activity

[1122] HOYER, Erlaubtes Risiko, 2009, p. 863.

[1123] HILGENDORF, Moderne Technik, 2015, p. 97.

[1124] FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 171 Rn. 12.

[1125] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 245; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 185 Rn. 58.

[1126] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 566.

[1127] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 448 Rn. 36.

[1128] SCHUSTER, Künstliche Intelligenz, 2020, p. 397 f.

[1129] VALERIUS, Sorgfaltspflichten, 2017, p. 10.

itself may be permitted, liability arising from traffic accidents caused by such activities is not encompassed within this permission, which leads to liability. However, given the nature of this concept, permission extends not only to the risk itself but also to the harm arising from it within the authorised framework[1130].

(2) Debates on the Legal Nature of Permissible Risk

The absence of a clearly defined legal norm explicitly addressing permissible risk -regardless of whether such a norm is necessary- makes the content, scope, and dogmatic position of permissible risk highly controversial, and in this regard, its legal nature is assessed within different categories[1131]. The debates extend to questioning whether the legal concept of permissible risk even exists[1132]. According to some, permissible risk is not based solely on a uniform principle, rather to various aspects of criminal law evaluations[1133]. The only point of consensus is that permissible risk does not give rise to criminal liability[1134].

Legal theorists have characterised permissible risk as a flexible concept, noting that it is difficult to define and apply through strict rules. Given this ambiguity, it must be applied with caution. Particularly, if the case involves *e.g.* a justification ground that eliminates the need to discuss the concept of permissible risk, that justification should be applied primarily[1135]. In this context, it has been argued that permissible risk is not an independent principle that justifies or limits criminal actions on its own; but is instead a formal term that indicates the presence of allowable risky actions based on

[1130] HILGENDORF, Moderne Technik, 2015, p. 99.
For a critique of this perspective, which also considers German legal dogmatics and argues that this view is logically flawed because what is permitted is the outcome that is violating legal interests; see: ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 359.

[1131] PREUß, Untersuchungen zum erlaubten Risiko, 1974, p. 227; MITSCH, Das erlaubte Risiko, 2018, p. 1162; HILGENDORF, Moderne Technik, 2015, p. 97 f.; GLESS, Mein Auto, 2016, p. 240; HEGER, StGB § 15 in StGB Kommentar, 2023, p. 44

[1132] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 405.

[1133] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 401.
For a comprehensive discussion, see: KIENAPFEL, Das erlaubte Risiko, 1966, p. 28 f.

[1134] GIEZEK, Einige Bemerkungen, 2009, pp. 545-546.

[1135] MITSCH, Das erlaubte Risiko, 2018, p. 1166.

various legal reasons. Despite being a formal concept, it plays a significant role in the legal system by grouping together cases where dangerous actions are not considered wrongful[1136].

Debates on the legal nature of permissible risk mainly focus on whether it constitutes a factor limiting the duty of care in cases of negligence, an approach that restricts the elements of the offence, a special justification, or a ground for excluding culpability. According to the adopted view, its scope of application is closely related to, and even overlaps with, other concepts such as social adequacy and objective imputation[1137]. It has thus been argued whether there is a need for a separate legal concept, such as permissible risk, to formally allow risky actions. Existing legal rules already permit risk-taking in various contexts. Creating a distinct category solely for risky actions may be redundant, as each case requires specific justifications for permitting the risk[1138].

According to the prevailing view, permissible risk serves to limit the required standard of care in cases of negligent liability and to refute unfounded accusations of negligence[1139]. In this regard, the doctrine of permissible risk, originally developed to exclude socially accepted yet dangerous activities from criminal liability, has evolved to address negligence by normativising the absence of due care and emphasising risks mitigated by safety precautions as a basis for excluding liability[1140]. Thus, the permissible risk doctrine is employed to assess whether the objective duty of care in cases of negligence has been breached. Accordingly, in a specific case, an

[1136] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 425.
See also: PREUß, Untersuchungen zum erlaubten Risiko, 1974, p. 227 f.

[1137] KIENAPFEL, Das erlaubte Risiko, 1966, pp. 22-28; HILGENDORF, Moderne Technik, 2015, p. 97 f; AKSOY RETORNAZ, Otonom Araçlar, 2021, p. 343; MAIWALD, Zur Leistungsfähigkeit, 1985, p. 405.

[1138] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 411.

[1139] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 144 f.; HILGENDORF, Dilemma-Probleme, 2018, p. 700; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 302 f. Rn. 35 f.; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 185 Rn. 58 f.; DUTTGE, Erlaubtes Risiko, 2010, p. 144.; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 319 Rn. 823; HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 168 f.; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 227 Rn. 41.
See also: ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1186 f. Rn. 10 ff.

[1140] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1159 f., Rn. 214 f.

individual who exceeds the boundaries of risk deemed acceptable within the flow of social life is considered to have acted in breach of the duty of care[1141]. Therefore, it is stated that the concept of permissible risk in the absence of negligent liability is only a concluding statement but does not eliminate the need for a detailed examination and does not provide a solution method[1142].

It is undeniable that the concept of permissible risk finds its most significant application in the field of negligent offences[1143]. Adhering to the duty of care generally ensures, although not invariably, that harm to others is avoided. Nonetheless, a residual risk remains alongside the duty of care, as it cannot be so strictly defined that every potential danger is eliminated. An overly cautious individual might reduce the risk of harm to almost zero, but this is not a standard expectation. Even a normally cautious person who causes harm despite acting in accordance with the duty of care remains unpunished, as such harm falls within the scope of permissible risk[1144]. Hence, those who do not exceed the standard that is generally accepted as permissible risk are not acting in a manner contrary to due care. In other words, permissible risk is nothing more than a formalised description of the degree of care that must be taken to avoid the perpetrator being accused of negligence[1145].

Nevertheless, it is argued that the concept of permissible risk may also be applicable in cases of intentional offences. Accordingly, there is no reason to limit this legal concept to negligent behaviour. Despite opposing views, the concepts of permissible risk and observance of due care can also be recognised as limiting not only negligent but also intentional offenses: if it is permissible to cause certain risks, this -in principle- must also apply to intentional behaviour, *i.e.* to all actions relevant under criminal law[1146]. However, this perspective has been criticised: permissible risk does not apply in intentional crimes because compliance with rules of care only exonerates one from the accusation of not having been sufficiently capable of acting. On the other hand, a person who is capable of avoiding harm

[1141] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 262 Rn. 21.

[1142] KIENAPFEL, Das erlaubte Risiko, 1966, p. 28.

[1143] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 46.

[1144] MITSCH, Das erlaubte Risiko, 2018, p. 1167.

[1145] MAIWALD, Zur Leistungsfähigkeit, 1985, pp. 409-412.

[1146] SCHAFFSTEIN, Soziale Adäquanz, 1960, p. 372 f.; STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011, p. 82 Rn. 32; HERZBERG, Vorsatz und erlaubtes Risiko, 1986, p. 7.

but still intentionally causes a result they recognise as probable always acts in breach of duty and therefore operates outside the scope of permissible risk[1147].

The perspective that examines the permissible risk doctrine within the framework of objective imputation is also quite prevalent. The elements that exclude the violation of the duty of care, as preferred by the prevailing opinion, correspond to those that negate objective imputation despite the realisation of an increased risk[1148]. It is widely accepted that, in practice, there is little significant difference between addressing this concept within the framework of objective imputation as the creation of unlawful risk or within the context of negligence as the lack of due care[1149].

According to the objective imputation theory, for criminal liability, the perpetrator must have created an impermissible risk, which subsequently materialised in the specific typical harm encompassed within the protective purpose of the norm. Even if the perpetrator has created a legally relevant risk, imputation is still excluded if the risk is permitted, and the outcome (resulting harm) cannot be imputed to the perpetrator. Therefore, the objective elements of the crime are not fulfilled, because the creation of an impermissible risk is a prerequisite for meeting the statutory definition of wrongdoing. On the other hand, it is not sufficient for liability that an individual exceeds the permissible level of risk by violating behavioural rules; additional assessments within the framework of objective imputation are also conducted[1150].

[1147] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 404 f.

[1148] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 575 f. Rn. 117, 129.

[1149] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1159 f., Rn. 214 f.

[1150] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 487 Rn. 65; MITSCH, Das erlaubte Risiko, 2018, p. 1167; HEGER, StGB § 15 in StGB Kommentar, 2023, p. 47, 52 ff.; KUDLICH, Objektive und subjektive, 2010, p. 684; HEINRICH, Strafrecht AT, 2022, p. 89 Rn. 245; RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 85 ff. Rn. 48-62; RÖNNAU, Grundwissen, 2011, p. 312; HOYER, Erlaubtes Risiko, 2009, p. 874.
For the view that permissible risk excludes the elements of the offence (*Tatbestand*), see: WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 824, Rn. 92.
For an evaluation, see: KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 186 Rn. 60; MITSCH, Das erlaubte Risiko, 2018, p. 1162.
For the views in Turkish legal literature, see: HAKERI, Ceza Hukuku, 2022, p. 188; AKBULUT, Ceza Hukuku, 2022, p. 258 f., 384.

The view delineating permissible risk through the objective imputation theory posits that this concept can be applied not only to negligent crimes but also to intentional crimes. However, one view posits that this approach confines the scope of permissible risk to crimes that require a specific result. It cannot be applied to abstract endangerment offences, as they lack a result, and therefore, there is no basis for the objective imputation of a result[1151].

In cases where the victim's own culpable behaviour contributes to the incident, there is no need to apply the concept of permissible risk, as objective imputation is already excluded[1152]. This principle may apply to individuals who misuse AI-driven autonomous systems in a faulty incorrect manner. In this case, manufacturers will be exempt from liability.

Another perspective explaining the legal nature of permissible risk asserts that it constitutes a ground for justification. Particularly in the classical doctrine, permissible risk was being evaluated within the context of unlawfulness[1153]. According to one view, permissible risk is a special form of the justification principle of overriding interest. In this context, presumed consent is considered a subcategory of this principle. Similarly, unavoidable erroneous assumptions regarding the factual conditions of a justification, as well as risky rescue operations, are also encompassed within this framework[1154]. For instance, Slovak criminal law is one of the few legal systems that explicitly stipulates permissible risk[1155], where it is argued that this concept constitutes a justification ground[1156].

[1151] MITSCH, Das erlaubte Risiko, 2018, p. 1162, 1167.

[1152] *Ibid*, p. 1167.

[1153] DEMIREL, Taksir, 2024, p. 255.
Explanations regarding *social adequacy* will be provided below.

[1154] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 262-273 Rn. 363 ff., 369 ff., 386, 417.
See also: HEGER, Vorbemerkung 4. Titel in StGB Kommentar, 2023, Rn. 29.
For the view that permissible risk can be classified as a material unlawfulness in terms of the distinction of material and formal unlawfulness, see: ZAFER, Ceza Hukuku, 2021, p. 379, 415

[1155] Slovak Penal Code explicitly regulates permissible risk as: Section 27 - **Admissible Risk**: "(1) An act otherwise criminal is not a criminal offence if someone, in accordance with the current state of knowledge, performs a socially beneficial activity in the area of production and research and if the socially beneficial result which is expected from the performed act, may not be achieved without the risk of jeopardising an interest protected by this Act. (2) Admissible risk shall not apply if the result to which such act leads is evidently **disproportionate** to the degree of risk or if the performance of the activity is clearly contrary to the generally

The opposing view argues that, even though the term suggests "permissible" (*erlaubtes*), it does not constitute a ground for justification. A justification serves as a permissive norm that legitimises the realisation of the entirety of the factual elements of an offence. If this were the case, the affected individual would be obligated to tolerate the harm and be unable to rely on justification grounds such as self-defence or necessity[1157]. Moreover, the concept of permissible risk does not have a separate application within the domain of unlawfulness and as a justification. The concept is unnecessary for justifying actions within the scope of unlawfulness, as existing justification grounds and legal frameworks already offer sufficient criteria for evaluating such cases. Therefore, legal practitioners do not need to mention or rely on permissible risk when analysing justifications like presumed consent, self-defence, or necessity[1158]. Furthermore, the prevailing view rejects the notion of permissible risk as a justification for negligent offences, arguing that it is logically inconsistent to both breach a duty of care and be justified by acting within the bounds of a permissible risk[1159].

Finally, while permissible risk's legal nature is assessed under various categories, it reveals its impact in limiting criminal liability when a violation of a legal interest has occurred. The critical question here remains unresolved: what are the substantive criteria for determining permissibility, and who defines them; the legislator or the criminal law practitioner[1160]?

It is evident that establishing the legal status of permissible risk requires a thorough investigation of the foundational theoretical aspects of criminal law dogmatics, given its complex interconnection with diverse legal frame-

binding legal regulation, public interest, principles of humanity, or it contravenes good morals."
Slovak Penal Code, 300/2005 Coll. ACT of 20 May 2005 PENAL CODE (as amended under Act No. 650/2005 Coll.), https://www.unodc.org/uploads/icsant/documents/Legislation/Slovakia/201124_CC_en.pdf.
See the original text: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2005/300. (accessed on 01.08.2025).

[1156] VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 669.

[1157] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 302 f. Rn. 35 f.

[1158] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 487 Rn. 65; MITSCH, Das erlaubte Risiko, 2018, p. 1167 f.
For instance, Walter does not classify sports competitions under the category of permissible risk and instead relies on the basis of full consent. See: WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 822, Rn. 90.

[1159] For the discussion, see: GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 587 Rn. 177.

[1160] MITSCH, Das erlaubte Risiko, 2018, p. 1162.

works. The present study, however, offers only a superficial analysis of the legal nature of the permissible risk doctrine to shed light on crimes involving AI-driven autonomous systems. As detailed above, the issues of negligent liability and the duty of care are particularly prominent regarding liability of person behind the machine for crimes involving AI-driven autonomous systems. In this context, identifying which activities are permitted and exempt from liability holds significance, particularly for mitigating the risks associated with emerging technologies through the required duty of care. Accordingly, without engaging in a further deeper analysis, the discussion in this study will focus on evaluating the limiting effect of permissible risk on the duty of care in this context.

(3) The Role of Permissible Risk in Limiting the Duty of Care

(a) Underlying Premise: Risks are Inevitable

It is a fundamental concept in risk perception that no human behaviour is entirely free of risks nor is any (technical) system without flaws. Every action performed by an individual carries the potential to infringe upon the legal interests of third parties. From the moment an individual leaves their home; even within the four walls of their own home, they are surrounded by numerous risks, both minor and significant. It can therefore be stated that life itself is inherently risky[1161].

Enhanced diligence and meticulous attention can serve to mitigate risks, diminishing both the probability and the magnitude of potential harm. Nevertheless, the complete elimination of all risks is unattainable, even in the most carefully conceived and executed behaviour[1162]. The complete abolition of the risks can only be accomplished by either abstaining from all action or imposing a comprehensive prohibition on all activities[1163].

In this regard, for the continuation and advancement of societal life, the acceptance of a certain level of risk is inevitable and essential. The argument is made that excessive caution can be more harmful than benefi-

[1161] SANDER/HÖLLERING, Strafrechtliche Verantwortlichkeit, 2017, p. 197; MITSCH, Das erlaubte Risiko, 2018, p. 1164; ZWICK, Risikoakzeptanz, 2020, p. 32; SCHÖMIG, Gefahren und Risiken, 2023, p. 209; ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 364.

[1162] GIEZEK, Einige Bemerkungen, 2009, p. 548.

[1163] *Ibid*, p. 545; DUTTGE, Erlaubtes Risiko, 2010, p. 138.

cial. This is because in 99 out of 100 cases, no harm is done, and overly cautious behaviour for the sake of the potential harm in just one instance undermines societal dynamics[1164]. Controversially, it can be argued that it is the certain degree of caution that ensures that nothing happens in 99 out of 100 cases. Nevertheless, efforts to eliminate risks entirely may obstruct the development of innovative technologies and discourage developers; ultimately impeding societal progress and transforming life into a museum-like world[1165].

All industrial activities, technical systems and products inherently involve risks. Even the most frequently used and reliable computer programmes can show critical security vulnerabilities[1166] and programming errors (bugs) are, by their very nature, objectively inevitable[1167]. Errors in mass productions are unavoidable, and it is technically impossible to guarantee that all products will be 100% safe. As long as products meet a basic standard of safety, higher quality expectations depend on consumer demands. Marketing entirely flawless products is simply unfeasible[1168].

In this context, the advent of emerging technologies such as artificial intelligence introduces a novel set of risks that are often challenging to anticipate or identify in advance. It can be stated with statistical certainty that the widespread use of such systems will eventually, in some instances, infringe upon individuals' legal interests, cause harm, result in injuries; and in the worst cases, even lead to fatalities[1169]. Advancements in this field, where risks remain uncertain, may constantly face the threat of negligent criminal liability, potentially discouraging developers[1170]. Nonetheless, the only way to absolutely eliminate the risks posed by such systems would be the imposition of a comprehensive ban[1171].

It is therefore imperative that legislation and social structures should not seek the complete elimination of risks, but rather the reduction and management of such risks to an acceptable level. For technologies such as

[1164] MITSCH, Das erlaubte Risiko, 2018, p. 1167.

[1165] WELZEL, Studien zum System, 1939, p. 516.

[1166] RAUE, Haftung, 2017, p. 1842, 1844.

[1167] SPINDLER, IT-Sicherheit, 2004, p. 3147.

[1168] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 179.

[1169] FRISTER, 10. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 127 Rn. 6; BECK, Die Diffusion, 2020, p. 46.

[1170] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 243; HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 28.

[1171] BECK, Die Diffusion, 2020, p. 47; BECK, Das Dilemma-Problem, 2017, p. 129; TURNER, Regulating AI, 2019, p. 121.

AI-driven systems, it is the responsibility of both individuals and manufacturers to fulfil their duties of care by taking reasonable precautions. Given the impossibility of eliminating all risks and the inevitability of a small residual risk despite extensive testing procedures, reducing these risks to an acceptable level is the most rational way to preserve the benefits of such systems[1172]. Thus, the fundamental question becomes which risks may be created without the activity being considered unlawful and a breach of due care[1173].

Consequently, it must be acknowledged that, even in the most carefully designed systems, risks cannot be completely eliminated. To reduce these risks to an acceptable level, persons behind the machine must exercise the required diligence. In the context of AI-driven autonomous systems, while potential harms may be foreseeable and theoretically avoidable by refraining from production, manufacturers are nonetheless obligated to exercise due care to make the product as safe as possible. This can be achieved for instance, *inter alia*, by adhering to established standards, implementing software updates, and addressing bug fixes, product observation and support after sales[1174].

(b) Mitigating Risks to Permissible Thresholds

Having identified the permissible risk doctrine as a framework for defining the boundaries of the duty of care, the examination of the obligations placed on the person behind the machine becomes more essential. In this context, the boundaries of the duty of care, as detailed above[1175], are aligned with the measures required to mitigate the risks inherent in the relevant activity[1176]. Given the premise that the risks of certain activities cannot be entirely eliminated, every effort must be made to reduce those risks to a socially tolerable and acceptable level. Nevertheless, the obligation to mitigate risks is not unlimited; in parallel with what is expressed in the boundaries

[1172] HILGENDORF, Autonome Systeme, 2018, p. 113.

[1173] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 560 f.; KLEINSCHMIDT/WAGNER, Technik autonomer Fahrzeuge, 2020, p. 27 Rn. 33 f.

[1174] HILGENDORF, Dilemma-Probleme, 2018, p. 700; HILGENDORF, Moderne Technik, 2015, p. 103 fn. 21; LOHMANN, Liability Issues, 2016, p. 337 f.; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 281.

[1175] See: Chapter 4, Section C(4): "The Scope and Boundaries of Duty of Care for the Person Behind the Machine".

[1176] HILGENDORF, Moderne Technik, 2015, p. 99.

of the duty of care, individuals are expected to take measures that are reasonable and practicable, avoiding the imposition of an unreasonably excessive burden. However, this is directly linked to the risk inherent in the activity, and individuals must continuously seek ways to achieve the intended purposes with reduced risks[1177].

Observing the due care required does not, by any means, always require refraining from any behaviour that could impair the ability to avoid the realisation of an offence. Rather, society relies on the taking of risks in various areas like traffic and medical research, so long as these risks are kept within socially acceptable limits by following the relevant safety norms and standards[1178]. For instance, in the operation of a chemical plant, even if all safety regulations and precautions are strictly adhered to, accidents resulting in death or injury may still occur. However, the legal system permits such operations to proceed within the framework of socially permissible risks[1179].

In a 1978 ruling[1180], the German Federal Constitutional Court (BVerfG) addressed the constitutionality of laws governing the licensing of nuclear power plants. The court recognised that certain risks can be tolerated when the societal benefits significantly outweigh potential dangers. Specifically, regarding nuclear power plants, the court ruled that residual risks are acceptable if, according to current scientific and technological standards, harmful events are practically impossible. While acknowledging that catastrophic accidents cannot be entirely ruled out, the court found it permissible to limit fundamental legal interests for the sake of broader societal benefits, provided the risks are minimized and any unavoidable uncertainties are accepted as socially adequate burdens shared by all citizens[1181].

In recognition of permissible risk, manufacturers are obligated to take all reasonable measures to minimise risks associated with their products. This includes the continuous monitoring of products after sale and the implementation of countermeasures, such as recalls, when necessary[1182].

[1177] HILGENDORF, Gefahr und Risiko, 2020, p. 24 f.; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 447 f. Rn. 33; MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 176; MURMANN, Zur Berücksichtigung, 2008, p. 140.

[1178] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 302 Rn. 33.

[1179] MERAKLI, Ceza Hukukunda Kusur, 2017, pp. 193-194.

[1180] Federal Constitutional Court (BVerfG), decision of 08.08.1978, Case No. 2 BvL 8/77, reported in BVerfGE V. 49, p. 143.

[1181] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 220.

[1182] HILGENDORF, Gefahr und Risiko, 2020, p. 26.

If harm cannot be fully eliminated, they must adopt all reasonable measures, follow advancements in science and technology, and minimise harm both quantitatively and qualitatively[1183]. This obligation extends beyond the product's launch to its post-sale lifecycle, as long as the measures are reasonable. Defined by the principle of reasonableness, permissible risk aligns with product liability standards. Since these obligations are dynamic, manufacturers must keep up with new knowledge in accordance with state of the art to avoid negligence. Particularly concerning AI-driven autonomous systems, the determination of which risks are permissible will be a process shaped by social negotiation, in parallel with the risk-based approach outlined below[1184]. In this process, case law will play a significant role[1185].

In the case of emerging technologies, there may be known risks as well as unknowns. Manufacturers are obligated to research and implement new findings that can identify and mitigate previously unknown risks[1186]; thus new methods to identify and mitigate such risks, 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[1187].

Further progress is driven by learning from adverse outcomes. It means that, as development occurs, both standards and the duty of care will expand accordingly. For instance, if an accident occurs due to an unforeseen or previously unknown situation, the cause is investigated and understood in order to prevent its recurrence. Consequently, this knowledge should be integrated into the duty of care in the future[1188]. For instance, in parallel with the explanations regarding the evolution of the duty of care in negligence[1189], it can be understood -although it is debatable- that there was

[1183] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 561 f.

[1184] See: Chapter 4, Section C(5)(b)(1): "Risk-Based Approach".

[1185] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 561 f.

[1186] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 9.

[1187] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 164; HILGENDORF, Automatisiertes Fahren und Strafrecht - der Aschaffenburger Fall, 2018, p. 69; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 223.

[1188] NISSENBAUM, Accountability in a Computerized Society, 1996, pp. 33-34.

[1189] See: Chapter 4, Section C(4)(b)(4): "The Evolution of Duty of Care Through New Techniques" and Chapter 4, Section C(4)(a)(2): "Learning from Mistakes and Hindsight Bias".

no system in place during the *Aschaffenburg incident* to detect the driver's heart attack and take control of the vehicle[1190]. Indeed, the public prosecutor involved in the case has reportedly noted that it could not be expected for all safety measures to be implemented in every vehicle[1191]. However, in line with the evolving dynamic duty of care, modern vehicles are now being equipped with systems that detect when a driver loses control, such as in cases of fainting. These systems attempt to alert the driver with visual and audible warnings, tighten and release the seatbelt, and bring the vehicle to a safer position.

In the early years of using (semi)autonomous driving systems, it can be expected that challenging driving manoeuvres, such as sharp turns, lane changes, and merging in narrow lanes may not always be correctly managed by the system. Additionally, other difficulties may arise between self-driving vehicles and human drivers[1192]. If these systems are to become widespread, the duty of care for manufacturers and operators will be significantly higher until they are widely adopted and no longer make basic errors, with a focus on reducing risks as much as possible[1193]. These systems should not be subject to rigid behavioural requirements that would impede their development, but this should not lead to comprehensive carelessness or to unacceptable risks for uninvolved parties[1194].

In cases where the dangers of a system are known but no methods to avoid them exist, the product, in principle, cannot be placed on the market. However, manufacturers should be afforded some discretion to adapt to evolving risk awareness and advancements in technology[1195]. All assessments regarding the scope and boundaries of the duty of care in negligence apply to the mitigating of risks to an acceptable level. Moreover, the application of permissible risk also depends on an individual's abilities and specialised knowledge, as individuals apply their own expertise and skills to

[1190] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 804.

[1191] For the information see: HILGENDORF, Automatisiertes Fahren und Strafrecht - der Aschaffenburger Fall, 2018, p. 67 f.

[1192] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 62.

[1193] HILGENDORF, Wer haftet für Roboter? Autonome Autos. In: Legal Tribune Online (LTO), 21.07.2014

[1194] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 447 Rn. 30.

[1195] SPINDLER, IT-Sicherheit, 2004, p. 3147.

their actions, and this effects the avoidability of the harm. The behavioural norm, therefore, does not solely address hypothetical situations[1196].

Engaging in highly risky actions can constitute a breach of the duty of care by itself; such as when a manufacturer releases an untested, unpredictable self-driving vehicle software update for use on public roads, resulting in an accident. However, such extreme cases are rare, as new technologies are usually tested in controlled environments in stages, with efforts made to reduce their risks to a socially acceptable level[1197]. 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[1198]. What remains are *residual risks*, which are considered permissible[1199].

(c) The Impact of Permissible Risk on Negligent Liability

According to the prevailing opinion, under the permissible risk doctrine where the required duty of care has been fully exercised, criminal liability does not arise for residual risks. This is because, when all safety rules are followed, the behaviour is deemed to be cautious, and taking risks is permissible, as the individual is not held liable for outcomes that could not be avoided despite adhering to the necessary precautions[1200]. In this context, the focus lies on whether the individual took all reasonable measures to minimise the risk and whether such actions yield social benefits that, in the view of the legal community, justify or outweigh the anticipated collateral harm[1201]. Thus, the concept of permissible risk functions by delineating the

[1196] SCHÜNEMANN, Über die objektive Zurechnung, 1999, p. 216 f.; OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 246; STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 309 f. Rn. 16.

[1197] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 175 f.

[1198] GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 429; KAIAFA-GBANDI, Artificial intelligence, 2020, p. 315.

[1199] KINDHÄUSER, Zum sog. ‘unerlaubten’ Risiko, 2010, p. 404.

[1200] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 302 Rn. 34, DELOGU, Modern, 1987, p. 116 f.

[1201] HILGENDORF, Gefahr und Risiko, 2020, p. 13.

scope of the duty of care, particularly in the context of technologies that offer societal benefits[1202].

Although such risks are permitted for their broader societal benefits, it remains essential to differentiate between damages resulting from human error and those arising in inherently risky environments, such as road traffic, where compliance with safety regulations determines liability for damages. Accordingly, if a harmful outcome could have been averted by adhering to the relevant safety regulations, the perpetrator cannot invoke the inability to prevent the accident as a valid defence[1203]. Furthermore, even within the scope of permissible risk, strict liability under civil law remains applicable[1204].

In cases involving drivers who were driving slowly and in accordance with relevant traffic rules, the drivers would still be considered to have acted within the scope of permissible risk if they caused injury to a pedestrian, even though they maintained full control of the vehicle. As a result, they would not bear criminal liability for the harm caused. Though controversial, it is stated that this holds true even if the driver anticipated, expected, or deemed it likely that a pedestrian might cross their path. The key criterion here is compliance with the rules and specifically maintaining a speed within the prescribed limits[1205]. In contrast, it is argued that no one would consider it permissible to kill a pedestrian merely because a traffic accident was unavoidable despite the utmost care being taken[1206]. This matter requires a legal-political decision, and the scope of the area which is

[1202] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 164; HOYER, Erlaubtes Risiko, 2009, p. 874; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 319 Rn. 823; MAIWALD, Zur Leistungsfähigkeit, 1985, p. 413. See also: Strafrechtliche Produktverantwortung für Softwarefehler bei autonomen Systemen, Info-Brief vom 05.11.2019, https://www.jura.uni-wuerzburg.de/fileadmin/0200-ma-netze-direkt/Infoblatt/Infobrief_Strafrechtliche_Produkthaftung.pdf. (accessed on 01.08.2025).
According to one perspective, based on the concept of risk, negligence (as subjective imputation) does not lie in "exceeding the permissible risk", but in its individual recognisability. See: DUTTGE, StGB § 15 MüKo, 2024, Rn. 107.
According to the objective imputation theory, the creation of a permissible risk cannot constitute an (objective) breach of the duty of care. See: RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 532 Rn. 14.

[1203] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 403 f.

[1204] SCHULZ, Verantwortlichkeit, 2015, p. 199.

[1205] MITSCH, Das erlaubte Risiko, 2018, p. 1164.

[1206] STRATENWERTH, Zur Individualisierung, 1985, p. 294.

free from criminal liability; where threats to life and bodily integrity are not penalised on the basis of permissible risk, should be extremely limited[1207].

Indeed, fatalities may occur as a result of the use of self-driving vehicles and, statistically, this is almost certain. However, if every possible measure was taken to minimise harm during the design and production of the collision avoidance systems, and if the legal system has permitted its use, then the benefits of this system -including its overall potential to reduce traffic fatalities- may justify its classification within the scope of permissible risk. In such cases, the manufacturer cannot be accused of negligence[1208]. Nevertheless, in order to arrive at this conclusion, it is essential that the society shows a willingness to accept the associated risks and that the potential benefits can be demonstrated to outweigh these risks. Moreover, this must be assessed on a case-by-case basis for each AI-driven autonomous system application.

In this regard, one perspective maintains that drones, in terms of the potential dangers they pose and the number of individuals affected, cannot be considered under permissible risk. Conversely, production robots, due to the limited number of individuals exposed to them and the adequacy of protective measures, may be regarded as falling within the scope of permissible risk. Nonetheless, this does not directly imply that negligence liability will arise for drone systems; the due care requirements of the persons involved must also be specifically considered[1209]. In the assessment of semi-autonomous vehicles, on the other hand, where the driver temporarily relinquishes control to the autopilot, it is crucial to clearly define the scope of the driver's duty of care. Additionally, it must be assessed whether the autopilot's unpredictable behaviour falls within the scope of permissible risk[1210].

According to one perspective, until AI-driven autonomous systems are recognised and assigned their own criminal liability, the damages and crimes caused by these systems must be tolerated under *de lege lata* in criminal law (even if this is not a satisfactory solution). In light of these considerations, a certain degree of impunity could be embraced, particularly due to the potential benefits of such technologies. Instead, it should suffice to address the matter under civil law liabilities[1211]. On the other

[1207] GLESS, Mein Auto, 2016, p. 242.

[1208] HILGENDORF, Moderne Technik, 2015, p. 110 f.

[1209] SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 417 ff.

[1210] GLESS, Mein Auto, 2016, pp. 248-249.

[1211] SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 420.

hand, the necessity of an action is not the sole criterion for determining its permissibility; what matters is the unavoidable nature of the risk associated with the legally accepted action. If the risk cannot be avoided without averting the action entirely, the action is permitted, with the level of avoidability decreasing in proportion to the importance and indispensability of the action, as seen in the case of emergency vehicles[1212].

Finally, it can be argued that, undoubtedly, in a world characterised by inherent risks, an individual's ability to live freely and benefit from contemporary advancements is contingent upon the toleration of these risks to a certain degree[1213]. However, acting within the permissible risk must not result in a situation where all due care requirements become obsolete and, as a consequence, no longer need to be observed[1214]. For instance, the general permission granted for a hazardous activity or enterprise is intended solely for the operation under specific conditions. It does not constitute a *carte blanche* for any crime that may arise within the scope of its activities[1215]. Indeed, *Welzel*, in 1939, highlighted the danger that sophisticated criminals might exploit the concept of permissible risk as a cover, cleverly disguising their malicious intentions while committing crimes. In such cases, where intent is present, they should be prosecuted for intentional crimes[1216]. Nevertheless, this approach is not limited to intentional crimes. It should not result in circumstances where those developing and utilising emerging technologies invoke the concept of permissible risk to evade their responsibility to exercise due care. In each particular instance, the courts must meticulously evaluate whether the activity in question falls within the permissible risk and whether the persons behind the machine have adequately fulfilled their duty of care as required.

(d) Does Permissible Risk Cover Atypical Risks of AI?

After establishing that the permissible risk doctrine does not provide a *carte blanche*[1217] and that only certain risks can be deemed permissible under

[1212] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 245.

[1213] ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 353.

[1214] SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 419.

[1215] GLESS/SEELMANN, Intelligente Agenten, 2016, p. 19; ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 358; MAIWALD, Zur Leistungsfähigkeit, 1985, p. 423.

[1216] WELZEL, Studien zum System, 1939, p. 520 fn. 41.

[1217] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 423.

strict conditions, the question arises of whether atypical risks can also be considered permissible. To illustrate with the examples provided; while a tiger[1218], attacking passers-by after being released from a zoo represents a typical risk, spreading an infectious disease would be considered atypical. Similarly, a self-driving vehicle causing an accident by making an incorrect lane change is a typical risk, whereas the vehicle's software hacking into an information system would be atypical. The question then arises: how should the boundary between typical and atypical risks be defined? For instance, is it typical for a large language model (LLM) chatbot to use offensive language towards a user? What about sharing personal data obtained from one user with others because of a malfunction? Or deceiving people to achieve its goals[1219]? Undoubtedly, determining whether a risk is typical requires experience-based data, which is not yet available for AI-driven autonomous systems[1220]. In this case, can any offence committed by a chatbot be considered within the scope of permissible risk?

In my view, 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[1221]. Instead, the matter is more closely associated with the considerations highlighted above concerning the boundaries of foreseeability and the complexities arising from atypical causal processes[1222]. However, it does not appear feasible to accept that every atypical risk necessarily results in an atypical causal process, particularly considering the ambiguity surrounding the distinction between typical and atypical risks. Indeed, even at this early stage in the development of such systems, it is conceivable that risks

[1218] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 582.

[1219] STANLEY Alyse, "OpenAI's new ChatGPT o1 model will try to escape if it thinks it'll be shut down - then lies about it", 07.12.2024, https://www.tomsguide.com/ai/openais-new-chatgpt-o1-model-will-try-to-escape-if-it-thinks-itll-be-shut-down-then-lies-about-it. (accessed on 01.08.2025).

[1220] CHANNON/MARSON, The Liability for Cybersecurity, 2021, p. 2.

[1221] According to the objective imputation theory, behaviours that are generally socially acceptable, commonly tolerated, falling within the scope of general life risks, or merely increasing risks in a legally insignificant manner, do not constitute a legally disapproved increase of a risk. See: RENGIER, § 13. Objektiver Tatbestand in Strafrecht AT, 2019, p. 87 Rn. 51.

According to one perspective, determining whether the use of AI-driven systems constitutes a legally relevant danger under the doctrine of objective imputation, and whether this danger materialises in the actual outcome, highlights the critical importance of permissible risk and the scope of social adequacy. See: SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 416.

[1222] See: Chapter 4, Section C(4)(a): "The Boundaries of Foreseeability".

which are considered highly unexpected might nevertheless constitute typical risks. 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. 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 acceptable risks.

It can be argued that established practice and extensive debate in literature on the application of permissible risk (or social adequacy)[1223] and consent in sport competitions can serve as a guiding framework in this context. There are sports regulations and established practices tailored to the specific type of sport in question. While these measures cannot entirely eliminate all risks, they are designed to mitigate the likelihood or severity of harm inherent in the sport[1224]. Conversely, these rules are primarily concerned with the orderly flow of the game and are not determinative of the boundaries within the context of criminal law[1225].

In sports competitions, anyone who complies with the rules of the game does not breach a duty of care, and therefore, cannot be held liable for negligence if an opponent is unintentionally injured[1226]. Indeed, sports activities are often enshrined as rights in constitutions. A legal system that encourages and permits such activities while simultaneously criminalising injuries or deaths that naturally arise from them would render the exercise of this right impractical[1227]. In this regard, if the misconduct is a typical manifestation of physical sport, criminal liability is excluded. However, if the act is intentional or occurs outside the game or during a break, the defences of social adequacy or presumed consent cannot be invoked[1228].

[1223] For the relationship between permissible risk and social adequacy (*soziale Adäquanz*),see: Chapter 4, Section C(5)(b)(1)(b): "The Relationship Between Social Adequacy and Permissible Risk".

[1224] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 49 f.; GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 274 Rn. 421.

[1225] ESCHELBACH, Gefährliche Handlungen, 2020, p. 152 f.

[1226] *Ibid*, p. 151 f.

[1227] MITSCH, Das erlaubte Risiko, 2018, p. 1166; ÖZOCAK, Spor Ceza Hukuku, 2024, p. 221.

[1228] ESER, Zur strafrechtlichen Verantwortlichkeit, 1978, p. 374; ESCHELBACH, Gefährliche Handlungen, 2020, p. 151 f.; HEGER, StGB § 15 in StGB Kommentar, 2023, p. 36.

On the other hand, in contact-intensive sports such as football, it is not uncommon for players to sustain significant injuries, which can sometimes even be career-ending. In such instances, if the incident occurs unintentionally within the context of the game, the typical outcome is a red card, and criminal proceedings are rare. Nevertheless, it is difficult to ascertain how such situations can be resolved through the concepts of consent or permissible risk[1229]. The inadequacy of substantive law in addressing these cases, with recourse instead to procedural solutions such as refraining from initiating criminal proceedings *ex officio* or failing to report the incident, is far from satisfactory[1230]. According to one perspective, objectively and heavily exceeding the rules of a sport does not necessarily imply that the boundaries of permissible risk have also been exceeded. The scope of permissible risk should remain broad, as the only way to entirely avoid injury in sports is either to opt for a low-risk activity or to abstain from participation altogether[1231].

In literature, it is generally acknowledged that permissible risk encompasses the common risks inherent in the game. However, for harmful actions that are foreseeable but violate the rules of the game, the consent of the affected party is additionally required. Indeed, according to one view, injuries arising from rule-compliant play are generally considered socially acceptable, eliminating the need for explicit individual consent. Conversely, minor negligent rule breaches cannot be justified by implied consent or considered socially adequate, whereas grossly negligent or intentional breaches are entirely unacceptable[1232]. On the other hand, it can still be argued that minor breaches may fall within the scope of permissible risk, while criminal negligence would arise only in cases involving dangerous, gross, or reckless breaches of the rules[1233]. The general risk framework accepted by the legal system should be in the interest of the broader public. This tolerance must be confined to cases where the rule infringement does not reach a level of risk that exceeds what can be generally tolerated. Beyond such extremes, it would imply that the legal system has abandoned its duty to protect individuals' life and limb[1234].

[1229] ESER, Zur strafrechtlichen Verantwortlichkeit, 1978, pp. 369-372.

[1230] ESCHELBACH, Gefährliche Handlungen, 2020, p. 152 f.

[1231] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 49 f.

[1232] ESER, Zur strafrechtlichen Verantwortlichkeit, 1978, p. 372 f.

[1233] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1188, Rn. 284.

[1234] ESER, Zur strafrechtlichen Verantwortlichkeit, 1978, p. 372 f.

Another perspective on the matter asserts that, while athletes consent to foreseeable risks in their sports activities, consent alone is insufficient for severe injuries due to the limited autonomy over one's physical integrity. In this regard, it is argued that, instead of relying on a permissible risk concept to complement the individual's consent, sports activities should be considered as a *sui generis* ground of justification under the notion of "acknowledged risk". According to this view, individuals engaging in certain sports must assume certain risks, even if they do not explicitly consent to them. Indeed, no one consents to risks that could end their athletic career or even cause their death; however, undertaking such risks is a necessity to participate in the sport. The scope of *acknowledged risk* is limited to the typical risks inherent in the specific sport. For instance, while the possibility of death may be a risk assumed in taekwondo, it would not apply in bowling if harm results from an opponent's actions unrelated to the game. Moreover, the harmful outcome must occur during a sporting activity conducted within the rules of the game. For example, striking an opponent after the bell rings in a boxing match, or using a glove containing concealed metal would fall outside the risks acknowledged within this framework[1235].

It can be argued that the concept of *acknowledged risk* ignores the permissible risk doctrine, but instead serves as a means to overcome the technical obstacles to consent, such as the prohibition against consenting to death. Additionally, while it achieves almost the same outcomes as the combination of permissible risk and individual consent, it does so by classifying the activity as legally justified in its entirety from a juridical perspective. Furthermore, while the concept implies that all inherent risks associated with a specific sport should be anticipated and acknowledged, it lacks a clear delineation between typical and atypical risks. For instance, this approach does not provide a clear answer either, for example, in a scenario where a tennis player suffers a brain haemorrhage after being struck on the head by a ball.

In this regard, it would also be appropriate to address the concept of presumed consent, which may be relevant to the discussion. Presumed consent refers to a unique justification based on the reasonable assumption of the affected party's hypothetical will[1236]. It is argued that this concept, rooted in the permissible risk doctrine, constitutes a unique ground of

[1235] ÖZOCAK, Spor Ceza Hukuku, 2024, pp. 223-229.

[1236] ROXIN, Über die mutmaßliche Einwilligung, 1974, p. 453.

justification. It also provides the most suitable explanation concerning the duty of care imposed on the party presuming consent to ascertain the true intent of the affected individual[1237]. This is because the person presuming consent assumes the risk that the act may ultimately not align with the actual will of the holder of the legal right[1238].

In light of the explanations and past scholarly debates on the legal background to sports, it can be stated that recognising atypical risks under the permissible risk doctrine or considering them socially adequate appears to be challenging. 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.

According to one view, it is possible to rely on the assumption that latent risks will not materialise despite compliance with regulations[1239]. However, while certain risks associated with AI-driven autonomous systems may be considered within the scope of permissible risk, it is not feasible to evaluate all risks in this context. Due to the significant impact of such systems, the scale of atypical risks can reach extraordinary levels. For instance, a mass malfunction of self-driving vehicles could severely disrupt an entire city's traffic system and even cause significant harm to individuals. Therefore, treating atypical risks as permissible risks merely because the necessary duty of care has been fulfilled would amount to a *carte blanche*. This issue will be examined in greater detail below within the risk-based approach, focusing on evaluations based on the magnitude of the risk.

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. For instance, in cases such as a chatbot insulting a user (although this may be characterised as a typical risk), users could be informed in advance about the existence of such a risk and choose to accept it. However, this approach would

[1237] ERMAN, Ceza Hukukunda, 2003, p. 149, 238.

[1238] RÖNNAU, Vor §§ 32 ff in LK, 2020, p. 230, Rn. 217.
For the situation where a person acting based on presumed consent has not carried out a sufficiently careful examination of its conditions, see: ROXIN, Über die mutmaßliche Einwilligung, 1974, p. 452 ff.

[1239] MITSCH, Das erlaubte Risiko, 2018, p. 1165.

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. Similarly, scenarios such as a child's hand getting trapped in the doors of a self-driving bus are not situations to which consent would reasonably be given.

In conclusion, 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. Consequently, the application of permissible risk must be significantly narrower until greater clarity is achieved on the risks.

b. Recognising Permissible Activities: Legal Criteria and Analysis

(1) Risk-Based Approach

(a) Determining the Appropriate Risk Approach

i. The Concept of Risk

In modern society, the advancement of new technologies introduces novel risks across various fields. As a result, contemporary law increasingly focuses on risk allocation, addressing the widespread and previously unrecognised potential of such risks[1240]. A comparable theoretical discourse emerged with the introduction of automobiles, where the power of engines replaced horses. Ultimately, these risks were accepted in favour of the benefits of general mobility[1241]. Similarly, AI-driven autonomous systems are now employed across a range of sectors, including healthcare, transportation, finance, and customer service, among others. These systems impact different groups of individuals in various ways, offering countless benefits while simultaneously introducing distinct risks. Therefore, a universal risk

[1240] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 165.
[1241] GLESS, Mein Auto, 2016, p. 231.

approach or a general categorisation of permissible risk is not feasible. It is essential to delineate the specific benefits and risks inherent to each field, thereby establishing a standard of care and defining the scope of permissible risk in accordance with the specific conditions and circumstances of the activity in question.

Adopting an effective risk-based approach necessitates a comprehensive understanding of the concept of risk. Since individuals generally do not wish to be subjected to harm or loss, society takes certain risks in pursuit of potential benefits. For instance, individuals who take on investment risks in financial markets seek to grow their wealth. Accordingly, any risk-based approach must assess both the adverse and beneficial outcomes of an activity[1242]. The creation of risks should be accepted only to the extent necessary to achieve the intended social benefit, while those exceeding this threshold are to be condemned[1243].

The assessment of a risk as socially tolerable is typically determined by weighing its social usefulness and benefits against the magnitude and probability of the harm it may cause[1244]. However, these two factors are insufficient for a comprehensive risk-based approach. 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 in the assessment[1245].

[1242] EBERS, Truly Risk-Based, 2024, p. 9.

[1243] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 172.

[1244] *E.g.*, see: SCHROEDER, Die Fahrlässigkeitsdelikte, 1979, p. 257.
For instance, the EU's AI Regulation defines risk as "the combination of the probability of an occurrence of harm and the severity of that harm" in Article 3(2). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (*Artificial Intelligence Regulation*), 12.07.2024, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689. (accessed on 01.08.2025).

[1245] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 171; SCHÖMIG, Gefahren und Risiken, 2023, p. 162 f., 195; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 451 Rn. 44.

ii. The Balance Between Risks and Societal Benefits

The willingness to assume risks against potential harms arises from the pursuit of the benefits associated with such actions. Despite the foreseeability and avoidability of a harmful outcome, negligence may be excluded when the risk-creating behaviour provides substantial benefits, making certain damages tolerable. This reasoning is primarily grounded in a cost-benefit analysis[1246]. The limits of permissible risk are determined by an abstract balancing of interests, comparing the benefits of undertaking the activity with those of avoiding the associated risks[1247]. However, not every objective justifies potential victims having to tolerate the endangerment of their legal interests[1248]. The creation of unnecessary or easily avoidable risks cannot be regarded as permissible and should not be afforded any form of privilege. The permissible risk doctrine applies only when the intended socially beneficial applications inevitably involve the creation of certain risks. Even in such cases, the responsible party is under a strict obligation to minimise these risks to the greatest extent possible[1249]. Any risk creation that goes beyond what is absolutely necessary remains negligent[1250]. Accordingly, the duty of care is determined by the level of potential risks and the feasibility of implementing necessary safety measures or precautions[1251].

The determination of which activities fall within the scope of permissible risk is a political decision and lies within the domain of the legislator. Prohibitions and permissions must be carefully balanced, particularly by taking into account the assessment of the interests at stake[1252]. For instance, rather than permitting the risk explicitly, the legislator may adopt a nuanced regulatory approach, stipulating that, while the risk may not be permitted, it is also not subject to criminal sanctions. For example, in cases of negligent damage to property, there may be no criminal liability, but civil liability for compensation would still arise. Furthermore, the legislator may also prohibit the undertaking of certain risks and impose sanctions for violations

[1246] HILGENDORF, Gefahr und Risiko, 2020, p. 24.

[1247] FRISTER, 10. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 128 Rn. 7; FELDLE, Notstandsalgorithmen, 2018, p. 89.

[1248] MURMANN, Zur Berücksichtigung, 2008, p. 134 f.

[1249] HILGENDORF, Gefahr und Risiko, 2020, p. 24.

[1250] HILGENDORF, Moderne Technik, 2015, p. 110.

[1251] SCHÜNEMANN, Moderne Tendenzen, 1975, p. 576.

[1252] DUTTGE, Erlaubtes Risiko, 2010, p. 138; MITSCH, Das erlaubte Risiko, 2018, p. 1164.

of such prohibitions through administrative penalties instead of criminal ones[1253]. Frameworks for permissible risk must be established to prevent legal uncertainty and developmental impediments in AI-driven systems, particularly with regard to defining thresholds for tolerable malfunctions. In such contexts, a critical dilemma arises: the need to safeguard societal safety while avoiding excessive restrictions that could hinder innovation and limit freedom of action[1254]. According to one perspective, this balancing should not rely on a weighing of interests akin to that employed in cases of necessity[1255], as such an approach would introduce a utilitarian framework into the permissible risk doctrine. This is particularly problematic in situations where human life is at stake[1256].

In the context of permissible risk, a significant issue arises when one party (or a segment of society) benefits from a particular activity or technology, while another, whose interests are infringed upon through exposure to it, suffers harm. The permissiveness of such risks must be grounded on a clear and well-defined basis, whether it stems from societal consensus, public interest, or another appropriate framework[1257]. 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[1258]. Conversely, if users or others knowingly and voluntarily accept the associated risks, the threshold for permissible risk may be correspondingly higher[1259].

iii. Calibrating the Duty of Care Through Risk Levels and Public Tolerance

Whether a particular activity falls within the scope of permissible risk should be assessed using a risk-based approach. This evaluation -as explained above- considers factors such as the level of the risk, the benefits it provides, and the extent to which necessary precautions can mitigate the risk effectively. The benefit's qualification depends on the value of

[1253] MITSCH, Das erlaubte Risiko, 2018, p. 1165.

[1254] SEUFERT, Wer fährt, 2022, p. 329; GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 436.

[1255] See: Chapter 4, Section E(2)(b): "The Balancing of Interests".

[1256] DUTTGE, Erlaubtes Risiko, 2010, p. 139.

[1257] DUTTGE, Erlaubtes Risiko, 2010, p. 140 f.

[1258] BECK, Die Diffusion, 2020, p. 47.

[1259] SEUFERT, Wer fährt, 2022, p. 329.

the legal interests, their significance for the community, public opinion, the likelihood of success, and available alternatives[1260]. Such a risk-based approach aligns the duties and obligations with the level of actual risk by prioritising and calibrating enforcement actions proportionally to the identified hazards[1261]. For this purpose, while methods for establishing risk classes from other fields may serve as a reference, they cannot be directly transposed into criminal law[1262]. Thus, establishing risk classes offer an advantage over pure diligence standards by not only indicating whether duties of care apply; but also determining their intensity and quality, thus avoiding intuitive errors such as overestimating new risks and preventing overly strict decisions by judges lacking technical expertise[1263].

In German criminal law, *Schünemann* introduced a scale to assess the relationship between the risk of an action and its intended purpose. This scale classifies actions into four categories: *luxury actions*, *socially common actions*, *socially beneficial actions*, and *socially essential actions*. Each category reflects the level of societal significance and permissiveness of the associated risk[1264].

According to *Schünemann*, the acceptability of risks is determined by the social significance and necessity of the activity in question. As the social importance of an activity increases, both the degree of acceptable risk and the need for corresponding safety measures also rise. This creates a delicate balance between ensuring individual safety and achieving collective benefits. For example, non-essential luxury activities (such as walking predator animals in public spaces) posing even minimal danger are deemed negligent unless they are made completely safe; the public is not expected to take any precautionary measures for such activities. Socially accepted (common) activities (such as walking a dog (pet) in public spaces), which are common and embraced by society, are permissible if they involve a low level of danger and standard safety measures are sufficient, with minor residual risks managed by individuals exercising ordinary caution. In the case of socially beneficial activities (such as motor-vehicle traffic) that provide significant advantages to society, but cannot eliminate all risks despite reasonable safety measures, a moderate residual risk is therefore "permissible", and society cannot be expected to mitigate these

[1260] SCHÖMIG, Gefahren und Risiken, 2023, p. 290.

[1261] EBERS, Truly Risk-Based, 2024, p. 4.

[1262] SCHÖMIG, Gefahren und Risiken, 2023, p. 286 f.

[1263] *Ibid*, p. 294.

[1264] SCHÜNEMANN, Moderne Tendenzen, 1975, p. 576.

risks through personal precautions. Finally, socially necessary (essential) activities (such as railroads) that involve substantial inherent dangers, are permissible if additional safety measures are either impossible or would make the activity impractical. Larger residual risks should be tolerated in the overriding interest of society, as long as strict safety rules are followed without hindering the operation's practicability[1265].

Building on *Schünemann's* risk assessment framework, *Schömig* proposes the establishment of four distinct risk classes to determine the duty of care in cases of negligence: 1- socially disapproved or useless activities, 2- socially common activities, 3- socially useful activities, 4- socially required activities. Determining these risk classes, the uncertainty and the level of risk (probability of occurrence, extent and magnitude of damage)[1266] as well as the benefit and purpose (goal) of the action can be taken into consideration. The extent of the damage can be assessed based on an abstract ranking of the affected legal interests. For instance, in the context of economic interests, the extent of damage is determined by the material, financial, or monetary value involved. For non-economic interests, the severity of the impairment and the potential reversibility of its consequences are often the determining factors[1267].

Fig. 1: Level of Risk[1268]*:*

Extent of Damage	**Probability of Occurrence**			
	Low	Medium	High	uncertain
Large	3	3	4	4
Medium	2	2	3	3
Low	1	2	3	2
Uncertain	2	3	4	Uncertain

Level of Risk: 1: Low Level * 2: Medium Level * 3: High Level * 4: Unacceptable Level.

[1265] *Ibid.*

[1266] The author suggests 5 different risk classes: low, medium, high, unacceptable and uncertain.

[1267] SCHÖMIG, Gefahren und Risiken, 2023, p. 288 ff.

[1268] The tables (Fig. 1 and Fig. 2) are based on Schömig's work and has been translated into English by the author of this study. See: SCHÖMIG, Gefahren und Risiken, 2023, p. 292.

Fig. 2: The Level of Duty of Care to be Applied:

Benefit -(Social Acceptance)[1269]	**Risk Level**				
	Low	Medium	High	Unacceptable	Uncertain
Socially Disapproved / Useless	2	3	4	4	4
Socially Common	1	2	3	4	4
Socially Useful	1	2	3	4	3/4
Socially Required	1	1	3	3/4	3
Uncertain	1	2	3	4	4

1: Low duties of care and only as much as reasonable
2: Regular duties of care, as much as possible
3: Increased duty of care, as much as possible
4: Prohibited, except if lowering is possible.

These risk levels can be aligned with corresponding duties of care. At the lowest risk level, only minimal duties of care are required, constrained by what is considered reasonable. If even minimal duties are deemed unreasonable, no specific care obligations may apply. At the second and third risk levels, the duties of care are limited by what is technically feasible, with a distinction made between normal and increased levels of care for the higher risk. For activities falling within the highest risk level, they should generally be avoided unless the risks can be mitigated by reducing either the likelihood or the severity of harm[1270].

Such a risk-based approach, in conjunction with a duty of care framework that aligns with risk classes and evaluates both societal benefit and tolerance, is both appropriate and well-founded. In any case, it is essential to approach the matter based on the specific circumstances of the situation. Many methods for assessing dangers and risks necessitate case-by-case evaluations, requiring the integration of scientific and normative criteria to develop transparent and reliable risk classifications[1271]. For example, distinctions should be made between sports categories based on factors such as the level of violence, the likelihood of exposure to harm, whether

[1269] The original table has been adopted in accordance with views advanced in this study by adjusting the levels of duty of care considering AI-driven autonomous systems' risks. See the original table for the initial levels.

[1270] SCHÖMIG, Gefahren und Risiken, 2023, p. 288 ff.

[1271] *Ibid*, p. 232.

these risks are inherent to the nature of the sport, and whether the activity involves professional competition or is purely recreational[1272]. In this context, according to *Schünemann*'s classification, sports activities can be regarded as socially common (customary) and useful (beneficial) actions and, accordingly, a standard of duty of care appropriate to the level of risk should be established[1273].

According to one perspective, it is pragmatically difficult to explain why sports involving life-threatening risks, such as boxing and car racing, are permitted. The legal system, unable to prohibit certain long-standing practices, acknowledges them under the guise of "historical legitimacy", and framing them as socially accepted activities grounded in general consensus, which classifies them as socially customary activities[1274].

The acceptability of risky activities is likely to increase when they confer significant societal benefits. Conversely, for products with a lower societal value, such as toys, the tolerance for risk should be correspondingly lower[1275]. Although it is proposed that the risk level of an activity can be determined based on its societal benefits[1276], it can be argued that the advantages of an activity may not alter its risk level but merely influence its societal acceptability, and consequently, determine the extent of the duty of care expected from individuals. For example, it is argued that inherently dangerous activities such as hunting, which provide no clear benefits (and are even entirely harmful), can be permitted only in exceptional circumstances and under strict safety measures, with careful consideration given to whether the risks can be effectively controlled[1277].

In this regard, *Schünemann*'s argument that there is no societal benefit in allowing a predator animal to be walked in public spaces, and that it is therefore unreasonable to expect the public to tolerate such a risk, can be extended to AI-driven autonomous systems. In the classification of AI-driven autonomous systems, *inter alia*, the benefits they provide to different social groups should also be considered. It is unreasonable to expect

[1272] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 35 f.

[1273] GIEZEK, Einige Bemerkungen, 2009, pp. 544-545.

[1274] *Ibid*, p. 551; JAKOBS, 7. Abschnitt - Strafrecht AT, 1991, p. 201 Rn. 36.

[1275] GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 436.

[1276] SCHÖMIG, Gefahren und Risiken, 2023, p. 290.

[1277] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1160 f., Rn. 217.
In my view, no distinct area of permissible risk should be established, nor should the society be expected to tolerate one, in connection with an activity that should be categorically prohibited.

societal tolerance for technologies that benefit only a particular group, even if the duty of care has been fulfilled to the fullest extent possible.

(b) The Relationship Between Social Adequacy and Permissible Risk

Having established that certain risky activities may be deemed acceptable due to their societal benefits, it would be prudent to examine the concept of *social adequacy* (*soziale Adäquanz)* before analysing the legal implications of society's willingness to accept such risks. In legal literature, the concepts of social adequacy and permissible risk are frequently used in close connection. This doctrine is often described as an attempt to align the criminal law system with social reality[1278]. Indeed, the acceptance of risks can, in some cases, be derived from certain legal rules, but in most cases, it is due to their social acceptance. This brings the two concepts into closer alignment, as in many instances, acceptance of risks is based on social adaptation over time[1279].

The concept of *social adequacy* is applicable not only in criminal law but also in other fields, such as labour law, for instance, in cases involving the private use of company internet[1280]. However, the legal nature and scope of social adequacy, as well as its relationship with other related concepts, remain subjects of debate and have yet to be definitively clarified[1281]. To illustrate with an example, it is stated that, when a car overtakes a motorcycle during lawful driving, there is always a possibility that the motorcycle might suddenly swerve and collide with the car. In this context, the situation of the car driver can be approached through *Binding*'s *permissible risk doctrine*, *Mezger* and *Blei's* notion of *relevance theory*, or *Welzel*'s *social adequacy theory* as well as within the framework of the *principle of reliance* or the modern theory of imputation[1282]. Nevertheless, the circumstances differ if it becomes evident that the motorcyclist is likely to make a sudden manoeuvre, similar to if it were apparent that an AI-driven system is susceptible to malfunction.

[1278] RÖNNAU, Grundwissen, 2011, p. 311.

[1279] HILGENDORF, Gefahr und Risiko, 2020, p. 25; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 265; MERAKLI, Ceza Hukukunda Kusur, 2017, p. 194, fn. 385; AKBULUT, Ceza Hukuku, 2022, p. 258.

[1280] RÖNNAU, Grundwissen, 2011, p. 311.

[1281] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 263-273 Rn. 369 ff, 386, 417.

[1282] For the evaluation, see: KAUFMANN, Objektive Zurechnung, 1985, p. 267.

Even *Welzel*, who originally conceptualised the theory, underwent a shift in his views regarding its legal nature over time. In his 1939 work, he characterised permissible risk as a specific subset of socially adequate behaviour, primarily distinguished by the degree of legal risk posed to protected legal interests. Activities falling within the scope of permissible risk are not subject to criminal sanctions due to their societal utility and the necessity of such risks[1283]. His approach to the elements of crime went through significant revisions in the more recent editions of his textbook, which, in turn, influenced his conceptualisation of *social adequacy*. Initially, he argued that the concept excluded the elements of the crime (*Tatbestand*), but later he re-evaluated this position, considering it within the framework of unlawfulness. Accordingly, social adequacy has been evaluated as a justification for behaviour based on the facts, rooted in the social-ethical order of community life. In this context, while he initially included intentional offences within the scope of his analysis, he later re-evaluated his argument and focused predominantly on negligent offences[1284].

The social adequacy theory posits that certain minor behaviours, which are deemed socially acceptable, are not subject to punishment due to their historical socio-ethical order of community life, tolerated within the society[1285]. For example, taking a few apples from the branches of a tree which extend over a public pathway[1286], or giving a gift to a postman on New Year's Eve are socially common behaviour and the latter would not constitute an offence under Section 331 of the StGB, which normally prohibits the acceptance of benefits[1287]. While the consensus among views on social adequacy is that such actions should not be punished; some scholars explain social adequacy as excluding the elements of an offence, while others describe it as a justification ground[1288].

The ambiguity surrounding the determination of the scope of social adequacy and the determination of behaviour deemed socially adequate has been criticised for leading to vague, inconsistent, and arbitrary refer-

[1283] WELZEL, Studien zum System, 1939, p. 518.

[1284] For the assessment, see: PETERS, Sozialadäquanz, 1974, p. 419. See also: SCHAFFSTEIN, Soziale Adäquanz, 1960, p. 373 fn. 11.

[1285] WELZEL, Das deutsche Strafrecht, 1969, p. 55 ff.; ROXIN/GRECO, § 10. Die Lehre vom Tatbestand in Strafrecht AT, 2020, p. 395, 398 Rn. 33, 40.

[1286] ZAFER, Ceza Hukuku, 2021, p. 379.

[1287] ROXIN/GRECO, § 10. Die Lehre vom Tatbestand in Strafrecht AT, 2020, p. 395 Rn. 33.

[1288] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 273 Rn. 418.

ences[1289]. The debate concerns the function of social adequacy insofar as customary activities are held to outweigh specific protective interests[1290]. Indeed, the lack of objective criteria for determining whether widespread practices in certain societies, such as male circumcision, fall within the scope of social adequacy creates ambiguity[1291]. A subjective perspective is even more problematic, as it risks encompassing highly objectionable practices such as female circumcision or even honour killings.

In this context, one perspective argues that, instead of relying on social adequacy that can lead to ambiguity, a restrictive interpretation based on the legal interest being protected offers a more accurate approach. This method avoids the risk of widespread abuses being excluded from criminal liability[1292]. A similar perspective holds that there is no actual need for a theory of social adequacy, as the same objective can be achieved through an interpretation consistent with the *ratio legis* of the norm[1293]. In contrast, another view contends that the criterion of whether the legal interest protected has been violated is itself prone to ambiguity. In fact, all proposed solutions to this issue inherently involve a degree of uncertainty; thus, the reliance on discretion and the assessment of judges in practice becomes necessary[1294].

Another related concept is the notion of insignificant acts. It is argued that refraining from penalising insignificant acts is grounded in their social adequacy and the lack of any violation of the legal interest protected by the norm[1295]. Due to the *ultima ratio* principle in criminal law, minor legal violations should not be subject to judicial punishment, necessitating a restrictive interpretation of the norm[1296]. The principle of refraining from penalising minor legal violations can also be derived from the constitutional principle of proportionality, which requires a balance between the offence and the punishment[1297]; a principle that must be observed not only by the legislator but also by the courts[1298]. Although certain legal systems

[1289] OTTO, Soziale Adäquanz, 2009, p. 226.

[1290] KINDHÄUSER, Zum sog. ‘unerlaubten’ Risiko, 2010, p. 408.

[1291] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 275 Rn. 427.

[1292] ROXIN/GRECO, § 10. Die Lehre vom Tatbestand in Strafrecht AT, 2020, p. 398 Rn. 41.

[1293] ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 356.

[1294] HAKERI, Ceza Hukukunda Önemsiz Hareketler, 2007, p. 85.

[1295] *Ibid*, p. 94 f.

[1296] *Ibid*, p. 63.

[1297] *Ibid*, p. 79.

[1298] ALBIN, “Sozialadäquanz”, 2011, p. 202.

include(d) provisions in their penal codes stating that insignificant acts[1299], even if they are typical (fulfil the elements of a crime), shall not be punished. It is argued that, without the need for such a provision[1300], it is more appropriate for judges to apply this principle through their interpretation in specific cases[1301]. In contrast, one view contends that insignificant acts in criminal law need not be addressed through social adequacy, as the same outcome can be achieved through a purpose-oriented interpretation that prioritises the protected legal interest[1302].

One view suggests that while determining the specific boundaries of the permissible risk area, the criterion of social adequacy should be applied, and decisions on whether a behaviour is permissible should be made based on its social usefulness or social acceptability[1303]. In contrast, another view distinguishes social adequacy from the concept of permissible risk. While it has previously been considered a justification or a basis for excluding guilt, the prevailing opinion today asserts that social adequacy serves to exclude the elements of the offence (*Tatbestand*)[1304]. A perspective that addresses the issue within the context of objective imputation argues that, due to their ambiguities and inadequacies, both social adequacy and permissible risk are unsuitable for example for the legal evaluation of sports injuries[1305].

According to a perspective with which I also concur, the concepts of social adequacy and permissible risk function on distinctly different conceptual levels. Social adequacy demonstrates that certain risky behaviours have been accepted by society over time on various grounds, and provides the substantive reasons rooted in societal norms for why an action is permissible. On the other hand, permissible risk indicates that a risky action is permitted under certain conditions without detailing the reasons. These concepts cannot be strictly delineated as they serve different functions within the legal system: permissible risk highlights allowable risks, whereas social adequacy explains the underlying reasons for permitting such risks[1306]. In other words, permissible risk is limited to referring to the per-

[1299] Such as the penal codes of DDR, the USSR, and Cuba. For the explanation, see: HAKERI, Ceza Hukukunda Önemsiz Hareketler, 2007, p. 67.

[1300] HIRSCH, Hauptprobleme, 1971, p. 140 f.

[1301] HAKERI, Ceza Hukukunda Önemsiz Hareketler, 2007, p. 93.

[1302] ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 122 ff.

[1303] For the evaluation, see: THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 284.

[1304] WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 823 f., Rn. 91.

[1305] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 52 ff.

[1306] MAIWALD, Zur Leistungsfähigkeit, 1985, pp. 408-409, 413.

missibility of certain risky actions, while social adequacy expresses factual reasons for the permissibility of certain actions[1307]. Indeed, general risks of life of normal magnitude have long been discussed under the concept of social adequacy. However, the social adequacy theory only serves as an interpretative tool rather than a method for determining which risks are acceptable[1308]. Accordingly, the elements of the offence should be interpreted in a manner that evaluates only socially inadequate conduct[1309].

(c) Society's Willingness to Tolerate Risks

For an activity to fall within the scope of permissible risk, fulfilling the duty of care to its fullest extent is not sufficient; it must also be established that the inherent risks are accepted by society. This societal tolerance is typically evaluated by balancing the activity's social utility and benefits against the level of risks involved. However, the question of how society accepts a given risk and how this acceptance can be determined remains essential.

Permissible risk can be understood as a collective-conventional agreement on the level of external hazard that society is willing to tolerate in exchange for certain benefits[1310]. Before deeming the risks of an activity permissible, it is essential, from the perspective of legal policy, to evaluate whether society is fundamentally prepared to accept even fatal accidents, as exemplified by those caused by self-driving vehicles. In such cases, criminal proceedings are likely to be rare[1311]. Determining which risks are deemed acceptable involves a process of social negotiation, where case law and legal debates will play a significant role[1312].

As discussed in detail, society accepts and utilises certain technologies, such as automobiles, despite their inherent risks (such as the risk of fatal

[1307] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 409.

[1308] STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011., p. 81 Rn. 30.

[1309] ROXIN/GRECO, § 10. Die Lehre vom Tatbestand in Strafrecht AT, 2020, p. 397 Rn. 37.
For a similar approach on interpretation of individual offences, see: KAUFMANN, Objektive Zurechnung, 1985, p. 268.

[1310] OGLAKCIOGLU, Strafrechtliche Risiken, 2023, p. 288.

[1311] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 573.

[1312] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 561-562.

accidents) due to the benefits they bring[1313]. While there is typically an inverse relationship between the level of risk and the extent to which society is willing to accept it, high-risk activities may still be tolerated if they offer substantial benefits. However, the willingness of society to tolerate such risks is not determined solely by the benefits they provide. It is influenced by a range of other factors as well. Some decisions regarding risks tend to be more intuitive than rational, particularly when fear plays a significant role in shaping perceptions and responses[1314].

Society's willingness to accept risks is influenced more by subjective factors than by rational calculations. Decisions in this context are not solely based on a cost-benefit analysis. These subjective factors can vary significantly between social groups. Key elements include the level of familiarity with the risk, the perception of control over the risk, and whether the risk was voluntarily chosen or imposed[1315]. Empirical research clearly demonstrates that risk-taking behaviour is significantly influenced by individual personality traits, social systems, situational conditions, and past experiences[1316]. Moreover, social communication plays a crucial role in shaping society's perception of risks[1317].

In everyday risk assessments, society tends to overestimate highly visible, rare, and human-induced risks (such as accidents, environmental diseases, and technological hazards) while underestimating systemic risks that develop gradually and are interconnected with positive developments (like climate change, resource scarcity, and economic imbalances)[1318]. Scientific risk assessments, which have advanced significantly, along with media communication, have ensured that many risks previously unknown to individuals are now widely recognised. For example, despite being statistically less dangerous than road traffic, fear of flying is widespread and considered to be risky, simply because people feel exposed in airplanes and the events are beyond their control[1319]. In this regard, social morality, with its diverse and often conflicting expressions in modern societies, influences the evaluation

[1313] GÜNSBERG, Automated Vehicles, 2022, p. 448.

[1314] SCHÖMIG, Gefahren und Risiken, 2023, p. 40.

[1315] HILGENDORF, Gefahr und Risiko, 2020, pp. 20-21.

[1316] LUHMANN, Ökologische Kommunikation, 2004, p. 136.

[1317] *Ibid*, p. 243.

[1318] SCHÖMIG, Gefahren und Risiken, 2023, p. 176 ff.

[1319] ZWICK, Risikoakzeptanz, 2020, p. 43, 49.

of new technological developments. Such advancements may be halted or rejected when moral judgments are codified into laws[1320].

Law is often shaped not by the rational calculation of risks but by the irrational social perceptions of individuals. Therefore, although societal acceptance of risks is sought for permissible risk, society's perception of risk -being highly subjective and susceptible to significant distortion- brings additional concerns and challenges. These critiques can also be directed at the concept of social adequacy. In this regard, it is argued that if risks were assessed based on objective criteria and guided by rationality, a new technology that likely causes less harm should be accepted[1321]. While this is a valid perspective, caution is required in risk assessment, as there is also the potential for society to irreversibly lose control over the technology. Furthermore, as will be examined below, it must be objectively demonstrated that the new technology brings less harm; however, due to the lack of empirical data, this determination is often challenging with new technologies.

Today, while simple examples of AI-driven systems are becoming an integral part of daily life, more complex ones, such as self-driving vehicles, remain largely absent from everyday use, particularly across much of the world. Undoubtedly, technical possibilities cannot be fully harnessed without risking harmful outcomes and potential criminal liability. It is widely accepted that the law can play a crucial role in facilitating these technologies by establishing specific duties of care and standards to manage risks. Once these technologies become normal phenomena of daily life, with their risks broadly accepted by society, and provided that the conditions set within the framework of duty of care are met, any remaining risks may be reduced to residual risks (yet it is still too early to consider these as the general risks of life). Achieving this requires the persons behind the machine to minimise risks through careful design and programming, rigorous testing, and continuous monitoring. Under such conditions, if the benefits of these technologies clearly outweigh their risks, the permissible risk doctrine may be applicable. Even though, this is not currently the case, over time, societal perceptions of risks evolve, and certain risks become increasingly acceptable. For example, society seems to be accepting the uncontrollable vast privacy violations that occur through smartphone use. Nevertheless, regardless of the social acceptance in the future, the persons

[1320] HILGENDORF, Modern Technology, 2017, p. 26.

[1321] HILGENDORF, Gefahr und Risiko, 2020, p. 22.

behind the machine could face criminal charges for avoidable design, manufacturing, or construction errors[1322].

In conclusion, five primary reservations regarding society's acceptance of the risks posed by new technologies can be noted. *First*, although society's perception of risk is inherently subjective, 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.

Second, the issue should not be assessed solely from the perspective of benefits outweighing risks; it is also crucial to consider the irreversible delegation of control from society to autonomous systems. As seen in the (near-future) use of autonomous taxis, the process begins with the delegation of specific tasks but is likely to evolve into the delegation of almost all activities in smart cities, leading to a significant diminution of human control.

Third, while emphasis is placed on society's acceptance of the risks and potential failures of AI-driven autonomous systems as a prerequisite for deeming such risks permissible, the question of how societal acceptance would manifest in scenarios such as the malfunction of military drone systems remains a matter requiring further discussion.

Fourth, it would be naive to suggest that this process unfolds within a framework of conscious and deliberate societal debate. In practice, fundamental rights and freedoms are often irreversibly altered through the interplay of rapid societal dynamics, advancing technology, and those who control it. A pertinent example is the swift abandonment of privacy concerns in the face of rapidly progressing technological developments.

Fifth, emerging technologies, such as smartphones, not only facilitate tasks previously undertaken by individuals but also gradually become new societal norms, thereby increasing the scope of personal responsibilities

[1322] GLESS, Mein Auto, 2016, p. 242; GLESS/WEIGEND, Intelligente Agenten, 2014, p. 587; GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, pp. 566-567, 573, 575; GÜNSBERG, Automated Vehicles, 2022, p. 448 f.
For a more sceptical view, see: WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 177, 229.
One view, for example, likens the mobility provided by self-driving vehicles for those who would not normally be able to drive to the opportunity glasses offer individuals with visual impairments to drive. In this regard, the required risk reduction capacity can be achieved through their proper utilisation. See: THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 29; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 295.

over time. In cost-benefit analyses, this phenomenon, which unfolds over time, is often overlooked.

(2) Assessing the Acceptability of Risks in AI-Driven Autonomous Systems

(a) Balancing Risks and Benefits

When evaluating whether a risk can be deemed permissible, it is crucial to consider objective criteria, such as the severity and extent of the potential harm, the probability and proximity of its occurrence, the ranking and value of the affected legal interests, the availability of prevention, mitigation and control measures, and whether the harm in question is irreversible[1323]. Following an examination of these factors, to determine whether society can tolerate the risks, the subsequent step is weighing the societal benefits of such activities against their potential dangers. This analysis constitutes another significant factor in determining the extent of the duty of care to be established in accordance with the aforementioned risk-based approach[1324].

The question of societal acceptance of risks for innovative technologies is not new and requires evaluation through the perspective of social usefulness, necessities and customs[1325]. A transparent societal debate is needed to assess where the benefits of AI-driven autonomous systems outweigh the risks and to define the boundaries of permissible risks[1326]. In evaluating the acceptability of risks, balancing society's various interests is crucial[1327]; 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. In light of the weighing up of different interests, overriding general interests often rationalise certain risks; for instance, road traffic serves as a prime example of a permissible risk[1328].

As examined in detail above, the prevailing approach in literature suggests that persons behind the machine must exercise all necessary care

[1323] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 171; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 451 Rn. 44; SCHÖMIG, Gefahren und Risiken, 2023, p. 162 f., 195.

[1324] See: Chapter 4, Section C(5)(b)(1): "Risk-Based Approach".

[1325] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, pp. 293-294.

[1326] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 451 Rn. 42.

[1327] LÜBBE, Erlaubtes Risiko, 1995, p. 960.

[1328] GLESS, Mein Auto, 2016, p. 240.

to minimise risks until such efforts reach a point where they become disproportionate. If further efforts to mitigate risks become excessively disproportionate or if certain risks cannot be reduced any further, it is envisaged that the remaining risks may be tolerated, because the probability of future damage cannot be excluded with absolute certainty. Manufacturers' assessment of user risks may be weighed against the broader burdens or implications of enhanced safety measures[1329]. However, from an economic perspective, it must always be remembered that mere efficiency gains do not justify higher accident rates; rather, the legitimacy and applicability of permissible risks depend on the enhanced safety provided by autonomous systems[1330].

In this assessment, the legal interest potentially infringed by the risk is of critical importance; for instance, in cases involving the potential violation of the right to life, the duty of care and the benefits necessitating the acceptance of such risks must be of the highest degree[1331]. In addition to the residual risks expressed in this manner, certain risks have been normalised due to their pervasive impact on societal life. For example, the fact that road traffic and its associated risks significantly shape the lives of individuals has led to the acceptance of these risks as a norm. Therefore, traffic risks are not regarded as residual risks but rather as general risks of life[1332].

On the other hand, due to the highly dynamic and complex nature of road traffic, it is difficult to classify the risks posed by autonomous vehicles as residual risks with today's technology. As technology advances, autonomous driving may become acceptable if risks are reduced below a certain threshold, provided they do not exceed current levels and are further reduced, particularly concerning life and physical integrity, given that autonomous vehicles will replace conventional cars[1333]. Indeed, society may be more willing to accept the risks of self-driving vehicles due to their benefits. However, the extent of such acceptance will be determined over time[1334].

[1329] Strafrechtliche Produktverantwortung für Softwarefehler bei autonomen Systemen, Info-Brief vom 05.11.2019, https://www.jura.uni-wuerzburg.de/fileadmin/0200-ma-netze-direkt/Infoblatt/Infobrief_Strafrechtliche_Produkthaftung.pdf. (accessed on 01.08.2025).

[1330] SANDHERR, Strafrechtliche Fragen, 2019, p. 4.

[1331] SCHULZ, Verantwortlichkeit, 2015, p. 193.

[1332] SCHULZ, Sicherheit im Straßenverkehr, 2017, p. 550 f.

[1333] *Ibid*, p. 551, 553.

[1334] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 176.

A recently published document by the OECD outlines the potential benefits and risks associated with AI while also presenting forward-looking policy recommendations[1335]. Nonetheless, AI-driven autonomous systems are employed across diverse domains and in various forms, making it impractical to conduct a universal risk-benefit analysis. In this context, the tailored application of the general risk-based approach outlined above; designed in accordance with the specific requirements of each case offers a prudent framework. This approach would effectively balance the interplay between the risk, scope of the duty of care, and societal acceptance[1336]. In this regard, for instance, autonomous systems developed for military purposes, self-driving vehicles, chatbots, voice assistants, and drones designed for entertainment each present distinct risks. The permissible risk thresholds for these systems must be determined based on the specific characteristics of the specific case at hand, by ensuring an appropriate balance with the corresponding societal benefits. Moreover, I contend that it is not feasible to establish a predefined *ex ante* permissible risk threshold for a particular activity or application. For instance, there is a significant difference between a chatbot exceptionally insulting a single user due to a failure and the same system simultaneously insulting all users (such as the *Gemini* incident, where it told a student “please die”, in contrast to *Grok*’s insulting thousands of users in July 2025).

Particularly regarding the unknown risks of new technologies, benefit-risk analysis must be conducted with greater sensitivity. A technical innovation can only be deemed legally permissible if it brings a substantial increase in benefits compared to the prior state of the art that clearly outweighs the additional risks it introduces[1337]. Furthermore, there may be unrecognisable risks arising from a lack of experience. If the persons behind the machine have fulfilled their duty of care by taking all conceivable precautions to minimise the danger, the question of whether society has accepted the associated risk is assessed. In such circumstances, if the benefits anticipated by society clearly outweigh the risks and disadvantages associated with the technology, it can be inferred that society is prepared

[1335] Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives, OECD Artificial Intelligence Papers, OECD Artificial Intelligence Papers No. 27, 14.11.2024, doi:10.1787/3f4e3dfb-en.

[1336] For the same view, see: HILGENDORF, Gefahr und Risiko, 2020, p. 17.

[1337] It is noted that the “substantially outweigh” test, as provided under Section 34 of the StGB, can be applied for this assessment: HOYER, Erlaubtes Risiko, 2009, p 880.

to tolerate these risks. Consequently, it is argued that an individual who suffers harm under such conditions is regarded as a victim of a risk collectively assumed by society[1338].

In the established literature, the applicability of the concept of permissible risk is assessed primarily on the basis of the benefits it yields for society. Accordingly, it is a logical inference that, in addition to considering such benefits, one must also take into account the harms and risks that both quantitatively and qualitatively diminish those benefits, as well as the drawbacks they generate from other perspectives. In this regard, before evaluating the social benefits and potential dangers of AI-driven autonomous systems, it is crucial to emphasise that such assessments must be conducted from multiple perspectives. For instance, what initially appears to be an advantage may simultaneously introduce significant risks and harms in the long term. To illustrate, although the use of robots and remote-control systems in armed conflicts might seem beneficial by reducing the resulting harm, including loss of human life; this could inadvertently diminish the motivation to avoid such conflicts. Consequently, attitudes towards armed conflict might shift, potentially leading to its more frequent occurrence[1339].

(b) Societal Gains of AI-Driven Autonomous Systems

It is evident that the societal benefits provided by AI-driven autonomous systems are the primary factor influencing their adoption by society. For example, despite concerns regarding the potential adverse effects of self-driving vehicles, including issues related to privacy and cybersecurity, a study involving 466 participants revealed that individuals recognised the potential of autonomous driving to significantly enhance road safety and efficiency. This finding suggests a willingness to balance perceived risks with the perceived benefits of technological advancement[1340].

Nevertheless, the assessment of (permissible) risk must vary across different AI applications. For example, in road traffic scenarios involving self-driving vehicles, society may be more willing to accept the associated risks, as the reduction in the frequency and severity of accidents benefits all road users. Conversely, in the case of medical devices equipped with

[1338] GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 435 f.
[1339] ANDERSON/WAXMAN, Law and Ethics, 2013, pp. 14-18
[1340] PRASETIO/NURLIYANA, Evaluating Perceived Safety, 2023, pp. 160-170.

AI systems, the risks are more likely to impact only those individuals who choose to utilise such technologies for their personal benefit[1341].

One of the most prominent applications of AI-driven autonomous systems, self-driving vehicles, aim to deliver several key benefits. These include enhanced road safety, improved mobility for individuals unable to drive, increased energy efficiency, reduced traffic congestion, and promotion of driver comfort and productivity[1342]. Indeed, the development of these technologies and the reduction of human involvement in road traffic are generally linked to improved safety. Although autonomous vehicles will not completely eliminate accidents or casualties, the common view is that they will significantly enhance safety[1343]. In this context, the Ethics Commission on Automated and Connected Driving, established by the German Federal Ministry of Transport and Digital Infrastructure, emphasised that "partially and fully automated traffic systems" are primarily designed to enhance the safety of all road users[1344]. Indeed, there are numerous instances where accidents that might have been unavoidable by human drivers have been successfully prevented through semi-autonomous driving technologies[1345].

In contrast to autonomous driving, human drivers may be subject to a number of potential limitations and impairments, including fatigue, distraction, and alcohol-related impairment. By eliminating these factors, autonomous driving can effectively reduce the likelihood and consequences of accidents caused by human error[1346]. Indeed, these systems never experience fatigue, intoxication, distraction from noisy environment, or the urge

[1341] OGLAKCIOGLU, Strafrechtliche Risiken, 2023, p. 289.

[1342] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 279.

[1343] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 452; SCHUSTER, Providerhaftung, 2017, p. 50 f.; DEUTSCHLE, Wer fährt, 2005, p. 252 ff.; THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 28.

[1344] Ethik-Kommission Automatisiertes und Vernetztes Fahren, Bericht der Ethik-Kommission Automatisiertes und Vernetztes Fahren, Bundesministerium für Verkehr und digitale Infrastruktur, June 2017, https://bmdv.bund.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile, p. 10. (accessed on 01.08.2025).

[1345] For some, see: "Top 10 Tesla Autopilot Saves", 30.08.2020, https://youtu.be/bUhFfunT2ds?t=45; https://www.youtube.com/shorts/eCLve-EJDGY; https://www.instagram.com/reel/DKo7V7uyQ9T/. See also: OWENS Jeremy C., "Driver in fatal Tesla crash previously had posted video of autopilot saving him", 01.07.2016, https://www.marketwatch.com/story/driver-in-fatal-tesla-crash-previously-had-posted-video-of-autopilot-saving-him-2016-06-30. (accessed on 01.08.2025).

[1346] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 447 Rn. 29; SCHULZ, Sicherheit im Straßenverkehr, 2017, p. 548.

to speed up to impress friends[1347]. In this context, various statistics indicate that 90% to 95% of accidents are caused by human error[1348]. According to the German Federal Statistical Office's 2011 statistics, 90% of all traffic accidents in Germany resulting in personal injury were caused by human error. General causes, such as weather, road conditions, and obstacles like wild animals on the road, accounted for 9% of accidents, while technical defects or maintenance deficiencies represented only 1% of the causes[1349]. Similarly, in the United States, 2015 statistics reveal that 94% of crashes were attributed to human choices or errors[1350]. Nonetheless, it is essential to note that the literature often reflects a misconception that such accidents would be entirely eliminated in the absence of human factors (*i.e.*, under autonomous driving).

With the widespread adoption of self-driving vehicles, the number of accidents caused by human error is expected to decrease dramatically[1351]. In this regard, it is argued that activating autopilot can be considered a permissible risk, and as self-driving vehicles become more prevalent, rare injuries may be regarded as general life risks[1352]. Furthermore, as the number of accidents declines, liability lawsuits will also decrease, offering economic advantages[1353]. On the other hand, while the overall number of accidents is expected to mitigate, it remains uncertain whether the severity of those accidents will increase or decrease[1354]. In particular, vehicles connected via a network are expected to experience fewer accidents quantitatively. However, self-driving vehicles may fail in circumstances where a careful human driver might avoid an accident altogether. Moreover, whether collision

[1347] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 294.

[1348] DEUTSCHLE, Wer fährt, 2005, p. 249; THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 28.

[1349] HÜTTER Andrea, "Verkehr auf einen Blick", Statistisches Bundesamt, Wiesbaden, 2013, https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Transport-Verkehr/Publikationen/Downloads-Querschnitt/broschuere-verkehr-blick-0080006139004.pdf?__blob=publicationFile, p. 39. (accessed on 01.08.2025). See also: LUTZ, Autonome Fahrzeuge, 2015, p. 120.

[1350] National Highway Traffic Safety Administration, "Federal Automated Vehicles Policy - Accelerating the Next Revolution In Roadway Safety", 2016, https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016 p. 5. (accessed on 01.08.2025). WAGNER, Produkthaftung für autonome Systeme, 2017, p. 709.

[1351] HILGENDORF, Teilautonome Fahrzeuge, 2015, pp. 16-17.

[1352] GLESS, Mein Auto, 2016, p. 233.

[1353] GOMILLE, Herstellerhaftung, 2016, p. 82.

[1354] DE CHIARA, et al., Car Accidents, 2021, p. 2.

avoidance systems can mitigate damage as effectively as human drivers, is not a straightforward question to answer. This issue requires further examination, particularly from the perspective of risk substitution, which is discussed from the perspective of substituting the risk, rather than merely decreasing[1355].

Another significant benefit of the widespread adoption of autonomous driving is the increased accessibility to individual mobility. This is especially beneficial for individuals with visual impairments, those who are too young or elderly, those with physical disabilities, or others unable to drive due to various circumstances[1356]. On the other hand, despite such advantages, if these individuals are legally and technically expected to intervene when necessary, the vehicles must be designed with simplicity and/or accompanied by appropriate training. This also requires the manufacturer to provide adequate information and fulfil necessary conditions. Still, if these individuals are unable to assume control of or operate the vehicle when necessary, respond to crucial warnings, or intervene in emergencies but still choose to use it, they may be held liable for negligent undertaking[1357].

Autonomous driving offers numerous additional gains in terms of environmental impact and efficiency. Particularly when integrated with networked vehicles, they offer significant benefits by improving traffic flow, reducing congestion, and lowering CO2 emissions. Through real-time data exchange, these systems can optimise road use, conserve resources, and enhance efficiency. Driver assistance technologies further contribute by automating monotonous tasks, increasing driving comfort. Additionally, innovations such as car-sharing and robo-taxis may enable more efficient, on-demand mobility, addressing individual needs while solving broader traffic challenges (such as the opportunity to adjust based on rush-hour conditions). By transforming road traffic into an intelligent network, autonomous vehicles promise time savings and a more sustainable approach to transportation[1358]. It is argued that self-driving vehicles, due to their sig-

[1355] See: Chapter 4, Section C(5)(b)(3)(a): “Substituting Existing Risks”.

[1356] HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 16 f.; THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 29; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 38, 64 ff.; FELDLE, Notstandsalgorithmen, 2018, p. 87.

[1357] See: Chapter 4, Section C(3)(d): “Negligent Undertaking”.

[1358] HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 16 f.; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 38, 64 ff.; DEUTSCHLE, Wer fährt, 2005, p. 252 ff.; FELDLE, Notstandsalgorithmen, 2018, p. 87; SCHUSTER, Providerhaftung, 2017, p. 50 f.

nificant potential, deserve more generous permissible risk standards than those applied to technologically simpler products[1359].

For example, a company's recently introduced *robo-taxis* promise numerous advantages, particularly in contributing to the sharing economy. It has been emphasised that a week consists of 168 hours, yet cars are typically used for only 10 to 15 hours, spending the rest of the time idle. As a result, traditional vehicles provide limited economic value to society[1360]. While this is a logical standpoint in many aspects, it overlooks the fact, as explained above, that such transformations occur as part of an interconnected system. Indeed, a sharing economy of this kind offers numerous potential advantages, but their full realisation depends on the system operating in a fully networked manner. In other words, these benefits can only be achieved if the envisioned future design entirely replaces the current framework. In this scenario, many aspects intrinsic to human life may become atypical, and even human drivers who opt not to use self-driving vehicles could be held liable for accidents. In my view, this is highly controversial, raising questions about whether this future truly represents a better society with greater overall benefits.

In this regard, it is important to emphasise that activities perceived to benefit society often alter various dynamics, and what initially appears advantageous may, from different perspectives or in the long term, lead to significant and unforeseen risks. For instance, chatbots like ChatGPT or Character.ai[1361] may offer educational or entertainment benefits; however, they could also risk aggravating problems by providing unproductive suggestions. Additionally, they might lead to further isolation from genuine human interaction and encourage laziness by discouraging individuals from actively researching and acquiring knowledge on their own. Determining the true impact is challenging, as it requires time and real-life experience, and it will likely involve a combination of both positive and negative outcomes.

[1359] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 585.

[1360] "Elon Musk Shows Off Tesla 'Robotaxi' That Drives Itself", 11.10.2024, https://www.nytimes.com/2024/10/10/business/tesla-robotaxi-elon-musk.html. (accessed on 01.08.2025).

[1361] An example of this is the case of a 14-year-old who became increasingly withdrawn and ultimately took their own life after forming a close bond with a character they had created on Character.ai. For the incident, see: ROOSE Kevin, "Can A.I. Be Blamed for a Teen's Suicide?", 23.10.2024, https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html. (accessed on 01.08.2025).

Beyond self-driving vehicles, AI-driven autonomous systems can perform tasks that humans cannot, prefer not to, or should not undertake (such as those that are dangerous, monotonous, or require high precision often executing them with greater efficiency and reliability than humans or traditional systems). For instance, autonomous systems are essential for modern space missions, particularly in deep-space exploration, where communication delays make real-time control from Earth impossible. Operating independently without continuous human oversight, these systems adapt to changing circumstances, learn over time, and incorporate user preferences, enabling more flexible and effective task execution. Moreover, when integrated into networks, autonomous systems can coordinate and collaborate with one another, enhancing overall performance and safety through collective action[1362].

The greatest benefits of AI (-driven) systems include cost reduction, quality improvement, and rapid response times[1363]. Additionally, they contribute intellectually by processing vast amounts of digital information and facilitating the integration of seemingly disconnected disciplines. Thus, they provide numerous benefits to society beyond the scope of this specific section, depending on its area of application[1364]. For instance, if an AI system analyses MRI images more effectively than a medical specialist but still has a margin of error, it can still be argued that its use would save more lives overall[1365]. In such cases, AI (-driven) systems should be utilised as decision-support systems combined with human judgement (human-in-the-loop) to further minimise risks. This approach aligns with the permissible risk doctrine, which requires the implementation of reasonable measures to mitigate risks. Since the evaluation focuses not on the technology itself but on the risks associated with the activity, the emphasis from a legal perspective here is on the risks of "AI outputs interpreted by humans".

While robots used in various fields can potentially cause physical harm to humans, advanced sensor and control systems enable them to proactively respond to human movements, significantly reducing the risk of injury. This capability is a crucial focus of research in physical human-robot interaction[1366]. In this regard, society will only accept a criminal law-free zone if harm to life and limb is minimised to the greatest extent possible.

[1362] SCHULZ, Verantwortlichkeit, 2015, p. 71 ff.

[1363] KIM, Implementation of AI, 2019, p. 144.

[1364] MÖKANDER/SCHROEDER, AI and Social Theory, 2022, p. 1349.

[1365] VALERIUS, Strafrechtliche Grenzen, 2022, p. 129.

[1366] ZECH, Risiken Digitaler Systeme, 2020, p. 26.

Achieving this requires systems to be designed to mitigate risks, allowing only those risks essential to achieving societal benefits. Any risks exceeding this threshold may be attributed to the manufacturer[1367].

(c) Potential Threats Posed by AI-Driven Autonomous Systems

It is evident that while AI-driven autonomous systems provide certain benefits, they also pose significant threats to different legal interests. Furthermore, the broader dynamics they alter often result in various harmful effects. A group of researchers from the *MIT AI Risk Repository* reviewed numerous studies and identified 43 AI risk classifications, frameworks and taxonomies, and compiled over 700 risks into a dynamic, continuously updated AI risk database[1368]. Indeed, a comprehensive examination of such risks goes far beyond the scope of this study. This section will briefly address key risks posed by AI-driven systems, including potential violations of fundamental rights and freedoms, network vulnerabilities, privacy threats, risks stemming from opacity, bias, loss of human control, degradation in the quality of generated outputs, unemployment, and energy-related challenges. These risks must be assessed in relation to the societal gains provided by the relevant activity to determine whether the associated risk can be deemed socially acceptable.

It must first be emphasised that objective and empirical data are essential for evaluating whether emerging technologies mitigate or exacerbate the existing risks associated with specific activities and pose other threats, thereby determining the acceptability of these risks. However, in the early stages of these technologies, there is a lack of sufficiently tested objective real-world data. Nevertheless, the permissible risk doctrine is of particular importance during the initial stages of technological development, where empirical data is insufficient. Thus, the initial challenges typically encountered during the early phases, where the precise nature and extent of the associated risks remain uncertain, create a paradoxical situation as to whether such risks should be permitted. This phenomenon which can be named the *develop-*

[1367] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 228.

[1368] SLATTERY Peter et. al., “The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence”, AGI - Artificial General Intelligence - Robotics - Safety & Alignment, V. 1, I. 1, 2024, doi:10.70777/agi.v1i1.10881, https://airisk.mit.edu. (accessed on 01.08.2025).

ment risk paradox bears similarities to the *Collingridge dilemma*[1369], yet it diverges by focusing on the dimensions of risk and their permissibility, with epistemic uncertainty lying at its core.

In this regard, while such systems have the potential to benefit society and are inherently desirable, one perspective holds that they should generally not be regarded as operating within the scope of permissible risk due to their inherent dangers and complexities, including issues such as opacity and autonomy risks. Exceptions should be assessed on a case-by-case basis, particularly in controlled environments where the associated risks are confined to a specific group of individuals. As a general rule, the greater the potential danger posed by an autonomous system, the less likely it is to qualify as a permissible risk[1370].

Even when initial "teething problems" of such new technologies are resolved and basic safety standards are met, new technological innovations often carry increased risks in their early market phase due to a lack of experience and incomplete testing for all possible real-world scenarios[1371]. While autopilots and similar AI-driven systems are highly effective in managing routine scenarios, they often struggle to navigate ambiguous or complex situations[1372]. For example, while autonomous driving is expected to reduce the overall number of accidents in the long term, individual accidents are almost certain to occur, with variations in their nature and form. Furthermore, current technology remains inadequate in effectively perceiving and processing challenging environmental conditions such as rain, snow, fog, dust, and significant fluctuations in lighting[1373]. Moreover, self-driving vehicles also cause accidents by committing basic errors that human drivers would be unlikely to make.

It can be argued that AI systems frequently fail to meet their grand promises made during their promotion, which are often designed to generate high expectations and persuade society to accept the associated risks. Despite hopes for fully autonomous vehicles, flawless medical diagnoses,

[1369] The *Collingridge dilemma* describes the challenge of regulating emerging technologies: early stages lack sufficient information for potential impacts, effective control and regulation; while later stages make changes difficult due to the technology's wide adaptation and entrenchment. See: COLLINGRIDGE, The Social Control, 1980, p. 19 f.

[1370] SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 419.

[1371] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 176.

[1372] GLESS, Mein Auto, 2016, p. 250.

[1373] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 66.

and perfect language processing, technical limitations and real-world complexity hinder AI's performance. Challenges like bias, transparency issues, and reliability gaps show that AI, while useful, cannot yet match the adaptability and nuanced understanding of human intelligence, especially in complex fields[1374]. This gap between expectation and reality highlights that AI (-driven) systems are tools, not flawless solutions.

The risks of bias and discriminatory outcomes associated with AI systems, arising from their reliance on historical data imbued with societal prejudices, constitute a significant concern. These biases have the potential to maintain unfair treatment in critical areas such as criminal justice, preventive policing, recruitment, and credit scoring; thereby aggravating existing social inequalities. Furthermore, the opacity inherent in many complex AI models hinders transparency and liability[1375], and ultimately undermines public trust in their application to delicate matters. To illustrate, an AI application utilising deep learning which exhibited gender bias led to erroneous results in diagnosing COVID-19 from medical images[1376]. Performing treatment based on such erroneous results without the supervision of a qualified human professional can lead to extremely detrimental consequences.

Autonomous systems driven by AI pose other significant risks due to the diminishing human control and oversight, which can lead to a reduction in human learning and decision-making capabilities, potentially resulting in disempowerment. Their physical presence and mobility increase the likelihood of physical harm to people and property, while their complexity and interconnectedness make them vulnerable to coordination failures and cyberattacks. In addition, these systems often collect and process large amounts of data unnoticeably, which raises serious privacy concerns, enable potential mass surveillance, undermine trust, and complicate legal liability due to a lack of transparency in their operations[1377].

One of the most critical and immediate risks posed by AI-driven autonomous systems is the threat of networking vulnerabilities. These systems increasingly operate as part of interconnected networks, communicating and coordinating with one another. With the expansion of 5G data transfer

[1374] Regarding AI's lack of reasoning, see: Chapter 3, Section B(2)(b): "Contra Arguments in Legal Literature Against AI-Personhood".

[1375] See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".

[1376] DERVISOGLU, et al., Unfairness of Deep Learning, 2021, p. 87 ff.

[1377] SCHULZ, Verantwortlichkeit, 2015, pp. 74-79. For an analysis concerning the risk of AI undermining democratic elections, see: BÖREKÇİ, Oy Hakkı, 2021, p. 632 ff.

capabilities, the proliferation of IoT devices, and the growing presence of self-driving vehicles; embodied AI-driven systems are becoming more prevalent and active[1378]. However, this interconnectedness significantly elevates the risk of cyberattacks. For example, in the context of smart cities, the risk of significant and widespread harm from the malicious exploitation of networked systems is a significant concern[1379]. While traditionally, a single vehicle or system might be compromised singly; the possibility of a mass-scale breach, *e.g.* through malware, poses far more severe threats. For example, stealing a conventional vehicle requires physical access, and an individual can typically control only one vehicle in this manner. In contrast, AI-driven autonomous systems connected to a network can be remotely hijacked and collectively manipulated or controlled, which significantly amplifies the associated risks[1380].

Networking risks associated with AI-driven systems can result in other swarm effects, leading to unforeseen and potentially devastating outcomes[1381]. Beyond the cybersecurity risks they pose; such vulnerabilities can open the door to other forms of exploitation[1382]. For instance, incorrect or biased learned conduct can quickly spread across interconnected networks, which amplify risks by rapidly implanting these flaws throughout entire systems[1383]. AI-driven systems can be manipulated for malicious purposes, and used to influence public opinion, spread misinformation, or affect elections[1384]. Additionally, hackers can exploit connected traffic systems to cause significant harm, such as steering truck convoys into small towns to create blockages, manipulating individual vehicles to accelerate or brake suddenly, or issuing faulty instructions that disrupt entire networks[1385]. Even robot vacuum cleaners could be easily hacked, allowing unauthorised

[1378] CHANNON/MARSON, The Liability for Cybersecurity, 2021, p. 17.

[1379] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 68.

[1380] HILGENDORF, Automatisiertes Fahren und Strafrecht - der Aschaffenburger Fall, 2018, p. 67; HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 154; VELLINGA, Cyber Security, 2023, p. 132 f.; CHANNON/MARSON, The Liability for Cybersecurity, 2021, p. 2.

[1381] ZECH, Zivilrechtliche Haftung, 2016, p. 175; ZECH, Risiken Digitaler Systeme, 2020, p. 27.

[1382] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 806.

[1383] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 450.

[1384] KATOĞLU/ALTUNKAŞ/KIZILIRMAK, Yapay Zekâ, 2025, *passim*.

[1385] SCHUSTER, Providerhaftung, 2017, p. 60.

access to their microphones and cameras. Such breaches can potentially lead to widespread privacy violations[1386].

Another significant risk associated with AI-driven systems is the potential for privacy violations. In a future where autonomous, networked sensors (in general data collectors) operate in a mass scale, the expectation of privacy is likely to diminish substantially (if there is any left). This erosion can occur in two primary ways: through the continuous collection of data by AI-integrated systems and the dependence on natural data to train and enhance AI systems. Particularly, the availability of natural data for AI development is increasingly limited, prompting a shift toward the use of synthetic data[1387]. As a result, natural (particularly personal) data, has become highly valuable and is frequently sought through both legal and illicit means.

It can further be argued that the delegation of numerous tasks to AI-driven systems may result in a reduction of human control, combined with an excessive reliance on AI. This may subsequently lead to a decrease in human oversight and an increase in moral and ethical uncertainties in areas where human judgement is essential. Consequently, this may elevate the risk of dehumanisation and erosion of the values which are essential to maintaining human-centred decision-making.

In my view, an additional factor that should be considered in the risk-benefit analysis of AI-driven autonomous systems is the potential for these systems to produce outputs of lower quality compared to those generated by meticulous human effort. At first glance, this issue may appear insignificant if these systems provide average-quality outputs while enhancing efficiency. However, the widespread reliance on such outputs could pose significant risks, particularly because newer AI models are often trained on the (average quality and synthetic) data generated by earlier models. An illustrative example is a legal professional who, with the assistance of such systems, might draft five documents in a day instead of one. While this apparent increase in productivity may seem beneficial, it raises concerns about a potential decline in the quality of the outputs, particularly in tasks requiring a high degree of precision and sensitivity. Over time, this degra-

[1386] In August 2024, security researcher, Dennis Giese, demonstrated at the Def Con Hacking Conference how *Ecovacs* robotic vacuum cleaners could be hacked: https://dontvacuum.me/talks/DEFCON32/DEFCON32_reveng_hacking_ecovacs_robots.pdf.

[1387] ZEWE Adam, "In machine learning, synthetic data can offer real performance improvements", 03.11.2022, https://news.mit.edu/2022/synthetic-data-ai-improvements-1103. (accessed on 01.08.2025).

dation could result in a feedback loop in which substandard data not only persists but also becomes increasingly embedded and magnified in such systems. To prevent such risks, it may be argued that human-in-the-loop mechanisms and oversight are necessary; however, in a system driven by efficiency, their implementation could become impractical.

Among the numerous risks associated with AI-driven systems, one of the most vital concerns that has generated significant public concern is the potential impact on employment. The potential displacement of human labour by these systems has been a subject of intense debate for many years. Beyond this, the environmental impact of AI presents another critical challenge. The training and functioning of AI models require significant energy, which has an adverse impact on environmental degradation.

Consequently, while numerous additional risks could be identified, they exceed the scope of this study. In my view, the primary focus in assessing society's willingness to accept a risk (and therefore permissible risk) should not merely be on whether a specific activity reduces risks or provides more gains in its immediate context. Rather, it is equally important to consider the broader dynamics it alters and the foreseeable effects of these changes in the near and medium term, in order to determine whether society can reasonably tolerate these risks. Indeed, while reducing risks in certain areas, such activities can simultaneously give rise to entirely new risks in others. For instance, while self-driving vehicles may generally reduce the likelihood of accidents, their widespread implementation could introduce systemic risks, such as large-scale malfunctions arising from network-related issues. Furthermore, in such evaluations, a new application that benefits one group may have adverse consequences for another. Legal systems must prioritise the public's utmost interests while striking a balance between competing legal interests.

(3) The Impact of Employing AI-Driven Autonomous Systems on Existing Risks

(a) Substituting Existing Risks

After examining the societal gains provided by AI-driven autonomous systems and their potential general dangers, it is essential to assess the impact of their use in a specific task on the existing level of risk associated with that task. For instance, when repetitive and monotonous tasks traditionally performed by humans are delegated to automated machines / systems, it can generally be argued that such a shift simplifies the process, offers numerous advantages, and even eliminates certain risks, such as injuries associated with these tasks, and makes automation a more preferable option. However, the situation may differ with autonomous systems. Rather than merely reducing specific risks, these systems might lead to a substitution of them. In other words, while mitigating some risks, they may simultaneously introduce new ones. Even in autonomous driving, one of the areas where AI-driven autonomous systems are claimed to offer the most benefit, this phenomenon can be observed.

From this perspective, technical innovations can broadly be classified into two fundamental categories. Firstly, risk-reducing innovations lower the level of risk compared to existing alternatives and can be generally deemed permissible without significant dispute. In contrast, other innovations which substitute risks offer enhanced advantages or utility but, simultaneously, introduce other (or higher) risks compared to existing systems. These innovations necessitate a careful evaluation, balancing the increased social utility they provide against the corresponding shift in risk[1388].

Without the need to examine complex systems, it becomes apparent that inventions presumed to fall into the first category, providing only benefits, may in fact introduce new types of risks. Seat belts and airbags used in automobiles serve as examples of this phenomenon. Indeed, while seat belts prevent serious injuries in the vast majority of accidents, they can, in certain cases, impede occupants from evacuating the vehicle and lead to fatalities. Similarly, airbags, despite their substantial benefits, may rarely deviate from their intended purpose, and pose risks such as suffocation and burns due to malfunctions[1389]. Nevertheless, due to the significant advantages they offer, their residual risks are legally accepted, provided that they adhere to the latest scientific and technological standards at the time of their introduction to the market[1390]. In this context, in 1979, the BGH highlighted that, while the use of seatbelts may present minimal risks (such as potential difficulties in rescue efforts after an accident) their benefits are overwhelmingly clear. Long-term data demonstrates that, for reasonable drivers, the advantages of seatbelt use far surpass these minor

[1388] HOYER, Erlaubtes Risiko, 2009, p. 878; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 224 f.

[1389] FELDLE, Notstandsalgorithmen, 2018, p. 90

[1390] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 222.

risks[1391], thereby indicating a strong and favourable benefit-risk balance[1392]. On the other hand, not all innovations substantially outweigh the existing risks they substitute. Even self-driving vehicles, which promise significant benefits and are expected to be safer than human-driven vehicles in the long term by reducing human error; introduce new risks such as hardware and software malfunctions, network vulnerabilities and potential hacker attacks, and unforeseen traffic scenarios, many of which have been detailed above; which results in a combination of reduced traditional risks and the introduction of new ones[1393].

Indeed, even today, numerous recorded accidents have been avoided thanks to the ability of semi-autonomous driving features to rapidly process environmental factors and execute manoeuvres. However, they have also caused fatal accidents by making fundamental errors that no human driver would ordinarily make[1394].

The reduction of risk in self-driving vehicles through collision avoidance systems, compared to human drivers, is a key condition for society to tolerate the risks associated with such technology. However, reducing the risk for one person may create risks for another. For instance, if a collision avoidance system prioritises the vehicle's occupants over pedestrians, while

[1391] Although this risk could lead to fatal outcomes, it has been classified as minor due to its low probability of occurrence.

[1392] Federal Court of Justice (BGH), judgment of 20.03.1979, Case No. VI ZR 152/78, reported in NJW 1979, p. 1363 ff.

[1393] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, pp. 221-225.

[1394] For some examples: "Tesla Autopilot feature was involved in 13 fatal crashes, US regulator says", 26.04.2024, https://www.theguardian.com/technology/2024/apr/26/tesla-autopilot-fatal-crash; "Tesla Full Self-Driving Drives THE WRONG WAY on ONE WAY Street in Downtown Atlanta", 07.10.2024, https://youtu.be/HVIvaYVfy5Y; The Wall Street Journal, The Hidden Autopilot Data That Reveals Why Teslas Crash, 13.12.2024, https://www.youtube.com/watch?v=mPUGh0qAqWA. Various dashboard camera recordings shared by users: https://x.com/missjilianne/status/1869565434481221879?s=12; https://x.com/thedooberhead/status/1869502131897782451?s=12; https://x.com/factschaser/status/1916623655129305491?s=12. See also: Paul Overberg, Emma Scott, Frank Matt, "Inside the WSJ's Investigation of Tesla's Autopilot Crash Risks", 31.07.2024, https://www.wsj.com/business/autos/tesla-autopilot-crash-investigation-997b0129; "Out-of-control Chinese AI car crashes into several cars - causing chaos on the roads", September 2024, https://telegrafi.com/en/Chinese-artificial-intelligence-car-out-of-control-crashes-into-several-cars-causing-chaos-on-the-road/. (Author's note for the last example: Despite extensive research, no additional sources could be found to confirm whether the accident truly occurred while the vehicle was in autopilot mode). (accessed on 01.08.2025).

this would be more advantageous for the occupants, individuals who walk to work daily would be exposed to a higher level of risk than before. As another example, if a vehicle suddenly brakes hard to avoid hitting a child who unexpectedly runs into the road, it could cause the vehicle to crash into a motorcycle following from behind, potentially resulting in fatal consequences for the motorcyclist[1395]. Such scenarios will be further analysed under dilemmas.

Another example can be drawn from the increasing use of e-scooters in urban areas. While e-scooters offer several significant benefits, such as facilitating individual transportation, contributing to the economy and environment through the sharing economy, and enhancing mobility; attention must also be paid to the risks associated with their use. For instance, if a person using an e-scooter is involved in an accident, the risk inherent in using this device becomes evident. By opting for the e-scooter -rather than a bicycle or car, which may offer alternative modes of transportation- the individual is substituting an existing risk, and this risk materialises when an accident occurs. In this example, it can be argued that a device which substitutes an existing risk, despite all its benefits, further increases the risk even when used in compliance with the rules.

Delegating a task to an AI-driven autonomous system similarly constitutes a substitution of risk. While this may reduce certain risks, it can simultaneously introduce new ones. The *vice versa* is also true: when a task is being performed by such systems and an individual intervenes to take over the task, this also results in a substitution of risk. For instance, in the event of an accident, if a driver of a semi-autonomous vehicle intervenes by recognising a hazardous situation and initiating an evasive manoeuvre, rather than allowing the system to respond autonomously, they must establish that their action was consistent with the duty of care. Alternatively, they must demonstrate that the accident would have occurred irrespective of their intervention[1396]. In any case, with regard to tasks delegated to AI-driven systems, if society is willing to accept the non-excludable residual risks associated with the use of such systems, given the overall benefits they provide (such as lower error rates and fewer accidents), then these risks may be regarded as permissible[1397].

[1395] OTTO, § 8 Pflichtbegrenzende Tatbestände in Grundkurs Strafrecht, 2004, p. 149 Rn. 202 ff.; FELDLE, Notstandsalgorithmen, 2018, p. 161.

[1396] GREGER, Haftungsfragen, 2018, p. 2.

[1397] VALERIUS, Strafrechtliche Grenzen, 2022, p. 129.

It can be argued that risk is not a quantitatively increasing or decreasing factor but rather one that varies in form depending on the specific circumstances of each case. In this context, another issue arises when new technologies simultaneously increase both risks and benefits or reduce one risk while increasing others. In such cases, the question becomes more complex, raising the issue of the extent to which risk should be permitted. One perspective suggests that if all those potentially at risk have been informed of the increased risk and have consented to it in pursuit of the additional benefits they seek, or have voluntarily exposed themselves to the risk, the permissibility of such risks could be rationalised[1398].

(b) Risk Enhancement through Task Delegation to AI-Driven Autonomous Systems: A Legal Analysis

When a criminal offence occurs as a result of a task being delegated to an AI-driven autonomous system, can the individual be held liable for having had the system perform the task instead of carrying it out in the conventional manner? Does the use of AI-driven autonomous systems increase the risk compared to alternative conventional methods? These questions are likely to arise frequently, particularly as such autonomous systems begin to replace traditional practices. To develop a legal solution in this context, it is necessary to examine whether the outcome would have still occurred even if traditional methods had been used instead of employing a robot for the task.

Various examples can be provided to illustrate the issue. For instance, a package might be delivered not through traditional means, such as by a regular vehicle, but instead by an autonomous drone. If the drone were to crash due to adverse weather conditions, causing injury to a person, this would constitute a relevant case for the analysis. Another example could involve a surgeon who, instead of performing a surgery manually, utilises AI-driven autonomous systems to assist in the procedure. If the use of such a system were to result in the patient's death, this would also represent a significant case for examination.

Undoubtedly, in such cases, the determination of negligent liability necessitates an examination of factors such as foreseeability, as outlined in detail above. Nonetheless, the primary focus here is on whether the use of

[1398] HOYER, Erlaubtes Risiko, 2009, p 879.

AI-driven autonomous systems has increased the risk of the specific activity and, consequently, whether the individual who delegated the task to such a system can therefore be held liable. In this context, it is essential to examine whether an alternative legally approved course of action would have also resulted in the same outcome and whether it increased the likelihood or severity of the harm, or endangered more serious legal interests.

This issue is frequently the subject of debate within the field of criminal law dogmatics. An example commonly cited in legal literature involves a truck driver overtaking a cyclist while maintaining a distance smaller than the legally required minimum. The cyclist, who swerves dangerously close to the truck, is subsequently run over and dies. It is later discovered that the cyclist was intoxicated, and it is certain that the accident would have occurred even if the truck driver had adhered to the legally required safe distance[1399]. Another example concerning AI-driven systems could involve a fatal accident caused by a fully autonomous vehicle. If the accident would have occurred even with the latest software update, which the owner or driver failed to install, could they still be held liable for the incident[1400]?

In such scenarios, determining the causal relationship between the breach of duty and the outcome can be challenging when the perpetrator has merely exceeded the permitted level of risk. It is widely accepted that in these cases, the perpetrator's specific breach of duty, namely, the legally disapproved danger created by their failure to exercise due care must have directly materialised in the specific outcome. While the perpetrator may have breached the duty of care, they were allowed to undertake the risk in question to a lesser extent[1401].

The perpetrator cannot be held liable if the outcome was objectively unavoidable. In other words, liability is excluded if the outcome would have occurred even if the legally approved risk-creating alternative behaviour had been conducted in compliance with the required duty of care[1402].

[1399] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 496 Rn. 88a. Kaspar argues that, in this case, the breach of duty and the dangerous situation created by it did not result in the cyclist's death, therefore, the truck driver cannot be held liable. See: KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 229 Rn. 50 ff.

[1400] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 174.

[1401] KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 304 Rn. 42.
See also: STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 311 Rn. 24.

[1402] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1129.

According to the prevailing opinion, it is not necessary to establish absolute certainty that the outcome would have been avoided if the alternative behaviour had been performed. Rather, if concrete indications suggest that the outcome might still have occurred even if the perpetrator had acted in accordance with the duty of care, the principle of *in dubio pro reo* applies. Thus, the perpetrator cannot be held liable[1403].

To illustrate, if it can be determined that the accident would have occurred even if the driver or owner had installed the software update, or that the patient would have died even with the ordinary surgical procedure, neither the driver nor the surgeon would be held liable. However, in addition to lack of experience on the matter, due to the opacity of AI, it may not always be possible to determine *ex post* why a particular outcome occurred[1404]. Unlike traditional systems, it may never be fully identifiable whether an alternative course of action would have prevented the harmful outcome. Nevertheless, according to the prevailing opinion on the matter, in cases where such a conclusion cannot be definitively determined, the principle of *in dubio pro reo* applies, and the perpetrator cannot be held liable.

The application of *in dubio pro reo*, despite an increased risk compared to alternative behaviour, has been criticised on the grounds that it excessively excludes dangerous acts from criminal liability for negligence[1405]. This is because the *raison d'être* of negligent offences lies in upholding duties of care, minimising risks as much as possible, and protecting potential victims[1406]. Indeed, in certain cases, an increase in risk compared to legally approved alternative behaviour may increase the chance of the occurrence of specific outcomes. While this cannot be definitively proven, conduct that increases risk beyond the permissible level, even if it has contributed to the

[1403] *Ibid*, Rn. 302 f., 1132; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 537 Rn. 35.
For an evaluation, see: KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 187 ff. Rn. 68 ff; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 305 Rn. 45 f; STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011., p. 84 Rn. 37; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 230 Rn. 54

[1404] See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".

[1405] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 496 Rn. 88b.
See also: KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 230 Rn. 55.

[1406] RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 536 Rn. 33 f.

occurrence of the outcome, remains unpunished. In this regard, according to the theory of risk enhancement[1407] (*Risikoerhöhungstheorie*) developed by *Roxin*, if an individual exceeds the legally permissible level of risk, any harmful outcome resulting from that increased risk becomes imputable to them[1408].

According to the theory of risk enhancement, whether there has been an increase in risk must be assessed *ex post*[1409]. If lawful alternative behaviour would certainly have led to the same outcome, the individual will not be held liable. However, if it cannot be definitively determined whether the outcome would have occurred, the result may be imputed to the perpetrator because they significantly increased the risk of the outcome compared to the lawful alternative behaviour. In conducting the analysis, attention is given to whether compliance with the permissible level of risk would have reduced the likelihood of the outcome and increased the chances of, for instance, the cyclist's survival[1410].

In this context, to avoid objectively imputing the outcome to the perpetrator, factors such as a decrease in the probability of the outcome occurring, a quantitative reduction in the extent of the damage, or the occurrence of a less severe result (e.g., bodily injury instead of death) are considered[1411]. On the other hand, it should be borne in mind that risk substitution may involve not only endangering previously unthreatened legal interests but also worsening the situation of an already threatened legal interest[1412]. However, if a person performs an act that has causal significance for the resulting outcome but does not in any way increase a pre-existing risk, and if the same outcome would have inevitably occurred even if that person had

[1407] As there is no established term for this concept in English legal literature, the term "theory of risk enhancement" has been adopted. Alternatively, the term "theory of increased risk" may also be used.

[1408] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 496 Rn. 88 ff.

[1409] *Ibid*, p. 499 Rn. 94.

[1410] *Ibid*; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 187 ff. Rn. 68 ff.; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 305 Rn. 45 f.; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 229 f. Rn. 53 ff.; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 302 f., 1132; GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 569 Rn. 86; ZIESCHANG, Strafrecht AT, 2023, p. 123 Rn. 435; FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 191 f. Rn. 81.

[1411] KINDHÄUSER/ZIMMERMANN, § 11 Objektive Zurechnung beim Erfolgsdelikt: Strafrecht AT, 2024, p. 103 Rn. 14.

[1412] STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011., p. 83 Rn. 35.

acted in compliance with the rules, they should not be held liable, even if an impermissible risk has been created[1413].

The prevailing opinion criticises the theory of risk enhancement for various reasons. First, it is argued that the theory merely ties criminal liability to the breach of the duty of care, thereby transforming (particularly negligent) criminal offences from breach of duty to endangerment offenses[1414]. Another objection to the theory of risk enhancement is that it violates the principle of *in dubio pro reo* in cases where it is not certain whether the outcome would have occurred regardless[1415]. It is further noted that all doctrines closely associated with objective imputation inevitably require a comprehensive balancing of goods and interests. Even those relying on standardised behavioural norms must acknowledge that such norms cannot eliminate the necessity for independent judicial assessment of the created risk. Additionally, placing excessive emphasis on risk enhancement unduly restricts the constitutional right to freedom of movement[1416].

In conclusion, it can be argued that delegating a task to AI-driven autonomous systems instead of using conventional methods may create new risks, increase existing ones, or allow the task to be carried out with reduced risk. Although some of these technologies are generally considered safer, during their early stages of adoption, they bring a range of unrecognisable risks. Therefore, despite being more resource-intensive, conventional methods should be preferred in cases involving significant legal interests such as surgeries (with the help of AI-driven systems if they will not increase risks unreasonably and the benefits balance such new risks). Increased efficiency, especially in situations involving significant legal interests, will not constitute a valid ground due to the potential for increased risk. 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 of the person behind the machine 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

[1413] ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 366.

[1414] RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 537 Rn. 35; ZIESCHANG, Strafrecht AT, 2023, p. 123 Rn. 435.

[1415] STRATENWERTH/KUHLEN, § 8 Die Tatbestandsmäßigkeit in Strafrecht AT, 2011., p. 84 Rn. 37.

[1416] DUTTGE, Zur Bestimmtheit, 2001, p. 127 ff.

outputs are often opaque and difficult to assess *ex post*. This could, in turn, incentivise unnecessarily “brave” conducts that excessively increase risk, effectively rewarding such conduct. In this regard, the arguments advanced by the theory of risk enhancement appear reasonable and should be taken into account, independently of whether the doctrine of objective imputation is adopted. However, this must not conflict with the adopted perspective of the legal nature of permissible risk.

(c) Does the Non-Use of AI-Driven Autonomous Systems Breach the Duty of Care?

When evaluating the impact of employing AI-driven autonomous systems instead of traditional and conventional methods on existing risks, an important consideration is whether the failure to utilise such systems might itself increase the risk and thereby give rise to liability for negligence. Indeed, if these systems become standard practice in the future due to their societal gains and especially their ability to mitigate risks, this matter will assume greater significance. In this context, it becomes essential to assess whether the non-utilisation of such systems amounts to a violation of the duty of care.

Many perspectives suggest that new technologies may become the new norm if they generally and essentially reduce risks. Particularly, if any new risks created by these systems are far outweighed by their benefits, and it is proven in the future that they pose significantly fewer risks, their use might even become mandatory[1417]. In such a scenario, if the maximum permissible risk is set to a level that can only be achieved through the use of the latest AI-driven autonomous technology, the failure to use these available systems could be considered a breach of the duty of care if it results in avoidable harm[1418]. For instance, it has been argued that several years after the widespread adoption and normalisation of AI-driven autonomous systems, such as self-driving vehicles, the use of regular vehicles could constitute a breach of the duty of care[1419].

[1417] GLESS, Mein Auto, 2016, p. 241.

[1418] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1122; CORNELIUS, Künstliche Intelligenz, 2020, p. 59; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 292.

[1419] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 179.

One perspective posits that when new technologies demonstrate a capacity to reduce risks compared to previous methods, they are considered within the scope of permissible risk. However, as a result, the older method, although potentially more profitable for the manufacturer may be deemed to fall within the category of impermissible risk. In such circumstances, the emphasis should be on prioritising the benefits to the general public rather than individual interests[1420]. The use of older methods should therefore only be allowed if the individual concerned provides informed consent. Any necessity to revert to the older method, particularly due to significant financial differences or similar considerations, requires that the individual be fully and explicitly informed of all associated risks[1421].

However, this perspective may be subject to criticism. Specifically, in cases where consent is absent or cannot be explicitly obtained -such as situations involving potential harm to the legal interests of a third party- it could effectively result in the total prohibition of older technologies[1422]. Indeed, particularly in the first few years of the transition from semi-autonomous to fully autonomous vehicles, there will be conflicts in the interaction between human and machine that will cause considerable damage. Particularly in smart cities where everything is interconnected through networks and is entirely designed around autonomous systems and self-driving vehicles, human drivers will probably become the atypical and unreliable element[1423].

Another issue may arise during interactions between machines. Particularly, compatibility problems can emerge between machines of different versions, expensive and inexpensive models, or older and newer technologies, due to disparities in performance classes. To enhance communication between machines, the legislature could establish certain performance catalogues, specifying the minimum requirements that machines must meet to ensure effective communication among themselves[1424]. On the other hand, although there is currently no legal norm mandating the use of autonomous driving systems[1425], it has been argued that prohibiting the use of older vehicles that are not sufficiently connected or equipped with autonomous

[1420] HOYER, Erlaubtes Risiko, 2009, pp. 878-879.
[1421] *Ibid*, 2009, p 879.
[1422] SANDER/HÖLLERING, Strafrechtliche Verantwortlichkeit, 2017, p. 200.
[1423] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 67.
[1424] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 560.
[1425] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 166.

functions would amount to a restriction of certain constitutional rights, such as the right to freedom of movement[1426].

Indeed, there are past examples where the deactivation of assistance systems has been deemed to constitute a breach of the duty of care. For instance, in an earlier court decision, involving a driver who deactivated the Electronic Stability Program (ESP) and subsequently forgot to reactivate it, it was determined that, had the ESP remained active, it was highly probable that the vehicle would have stayed within its lane. Therefore, the court not only regarded the driver's behaviour as careless but also classified the deliberate deactivation of the ESP as gross negligence under civil law[1427]. Similarly, in a decision by the German Federal Court of Justice (BGH)[1428], the non-use of a modern medical device was held to constitute negligence[1429].

Scenarios involving a human-machine combination require separate consideration. For instance, there are promising AI systems available today that can successfully detect cancerous cells more effectively than humans. However, these systems are not immune to errors and may produce false diagnoses[1430]. Therefore, instead of relying solely on their results to initiate treatment, the outcomes should be supported through additional testing to achieve the best possible result. Consequently, in human-in-the-loop activities like these, the new standard of care does not rely exclusively on traditional methods or solely on the new technology. Rather, it is the combination of the two that yields the optimal result. Any deviation from this approach would constitute a breach of the duty of care. To illustrate, in addition to numerous previous examples, in 2020, an African American man was wrongfully arrested by police in the United States after a facial recognition system misidentified him as a suspect. Despite his protests, the officers relied solely on the AI's identification[1431]. This incident underscores the necessity for humans to exercise caution and avoid overreliance on the

[1426] HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 22.

[1427] For the information, see: WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 167

[1428] Federal Court of Justice (BGH), judgment of 30.05.1989, Case No. VI ZR 200/88, reported in NJW 1989, p. 2321 f.

[1429] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 167.

[1430] CORNELIUS, Künstliche Intelligenz, 2020, p. 60.

[1431] RYAN-MOSLEY Tate, "The new lawsuit that shows facial recognition is officially a civil rights issue", 14.04.2021, https://www.technologyreview.com/2021/04/14/1022676/robert-williams-facial-recognition-lawsuit-aclu-detroit-police/. (accessed on 01.08.2025).

outputs of AI systems. This issue is particularly significant in the contexts of predictive policing, border control, and profiling.

A similar perspective arises in the context of autonomous driving, particularly regarding the possibility of vehicle malfunction. If it is proven in the future that self-driving vehicles are safer and result in fewer accidents compared to human control, overriding a properly functioning autonomous system could be classified as a breach of the duty of care[1432]. However, this scenario may create a dilemma in certain cases. From a general standpoint, if an occupant, who is expected to trust the vehicle (which is safer), intervenes due to a suspected malfunction and thereby causes an accident, the question arises whether the accident would have occurred regardless of the intervention[1433]. If an accident occurs in a scenario where the individual refrains from intervening, their liability for failing to act may be questioned. Setting aside the *ex post* issue of whether an alternative course of action would have altered the outcome, one view holds that penalising the individual in either scenario -whether for intervening or for failing to intervene- violates the principle of culpability[1434].

(d) Delegating Tasks to AI-Driven Autonomous Systems: An Alternative Approach for Liability

Autonomous systems driven by AI are progressively assuming tasks traditionally performed by humans[1435]. For example, driving is increasingly being delegated to vehicles with varying levels of autonomy, supported by continuously advancing systems. As discussed above, in the smart cities of the future, a significant portion of road traffic could consist of self-driving vehicles. In such a scenario, these vehicles might not even feature steering wheels or pedals. Human drivers could become atypical and might even be considered a luxury, potentially no longer regarded as a permissible risk.

As of mid-2025, a transitional period is proceeding. Tasks delegated to AI-driven autonomous systems are not limited to driving; gradual delegation is occurring across a wide range of fields, from household tasks to cognitive activities. While some of these tasks are partially delegated

[1432] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 167.
[1433] See: Chapter 4, Section C(5)(b)(3)(b): “Risk Enhancement through Task Delegation to AI-Driven Autonomous Systems: A Legal Analysis”.
[1434] THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 28.
[1435] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 547.

and remain under human supervision, others involve significantly reduced human oversight; in most cases, however, such oversight is steadily diminishing. The question of what humans would do if all tasks were delegated to machines falls outside the scope of this study. Rather, the emphasis here is on the delegation of inherently risky tasks, previously performed by humans to AI-driven autonomous systems, resulting in a gradual diminishment of human control. Consequently, humans gradually assume passive roles with corresponding reductions in their responsibilities and liabilities. Delegating a task in this manner can be likened, as observed in the literature[1436], to the practice of employing an individual of another faith to press the elevator button in adherence to the prohibition against using elevators on the *Sabbath*[1437].

As discussed earlier, the prevailing perspective in the literature suggests that the advancement of self-driving vehicles is leading to a shift in liability and control from drivers to manufacturers[1438]. This is largely accurate. However, caution is required against the assumption that drivers will transition entirely into the role of passengers with no remaining responsibilities. Such an analysis should not be limited to driving alone but should also consider the broader societal implications of diminishing control and the increasingly passive roles humans assume across various fields. In particular, it would be problematic to interpret this as a means of evading responsibility (and liability) by delegating the risks of an activity to systems that bear no criminal liability of their own[1439].

Nevertheless, contrary to the widespread opinion, I suggest 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. Legal systems should approach such situations cautiously and refrain from gener-

[1436] JOERDEN, Zur strafrechtlichen, 2020, p. 287.

[1437] KATZ Leo, Ill-Gotten Gains: Evasion, Blackmail, Fraud, and Kindred Puzzles of the Law, The University of Chicago Press, 1996, p. 24 ff.

[1438] See: Chapter 3, Section C(1)(d)(2): "Responsibility Shifting to Manufacturers".

[1439] This statement does not imply that such systems should bear criminal liability. See: Chapter 3, Section B: "Autonomous System's Own Liability".

alising that “autonomous driving will generally result in fewer fatalities”. Unless the individuals are entirely passive throughout the whole process, this point of activation or delegation of a task should form the basis for liability analysis. Nonetheless, 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. This issue is likely to become even more significant in the future as more tasks are delegated to AI-driven autonomous systems. 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 or bots despite their inherent risks. Whether such delegation falls within the scope of permissible risk must separately be evaluated.

Indeed, similar to the tiger released from the zoo[1440], the unpredictability of AI-driven autonomous systems is recognisable. Therefore, the argument of evading responsibility and liability by claiming that such risks are unforeseeable should be approached with caution. Delegating a task to a system that inherently involves low, medium, or high levels of risk constitutes an act of risk substitution. Accordingly, it is inaccurate to assert that such risks are entirely uncontrollable or unforeseeable. The moment of delegating control over the relevant task to these systems should serve as a starting point for liability analysis. Naturally, factors such as whether the conditions for negligence are met must also be carefully evaluated to determine liability.

Moreover, today individuals can still choose to delegate a task, whether currently performed manually or through automated means, to autonomous systems. In the future, however, most of the tasks will probably be performed by autonomous systems by default. In such cases, identifying the exact moment of delegation will often be unachievable. Liability analysis may only be feasible when a task is delegated to a system that is either riskier or safer than the default option. Ultimately, delegating a task to an autonomous system is foreseeable to involve varying levels of risk, and individuals who are aware of these risks must bear the responsibility for delegating their tasks by activating such systems.

[1440] GLESS/WEIGEND, Intelligente Agenten, 2014, p. 582.

In addition to the view that responsibility in self-driving vehicles shifts from the driver to the manufacturer, thereby absolving the driver of liability, there are further opposing perspectives on the matter. It has been stated that if there is no breach of the duty of care on the part of the driver; it would be incorrect to consider the activation of the system as constituting a breach of duty of care, as this would effectively amount to a prohibition on automated vehicles[1441]. Additionally, in the case of full autonomy, if the legislator decides to permit fully autonomous driving, the driver will no longer be held liable under civil or criminal law[1442].

Conversely, although such strict arguments have not been made, particularly regarding fully autonomous systems, similar views also exist. Accordingly, in the case of self-driving vehicles, where no driving action is performed by the user, the act of setting the appropriately programmed vehicle in motion becomes the starting point for criminal assessment[1443]. If the user decides to activate an autonomous system and can foresee the risks and harmful outcomes it may produce, their liability can be established[1444]. Indeed, delegating tasks to autonomous vehicles does not create a new sphere of responsibility, potentially leaving victims and society without anyone to hold accountable for the violation of their rights or interests[1445].

c. The Feasibility of Defining Permissible Risk Through Standards and Other Norms of Conduct

(1) Concretising Legal Expectations

In emerging technologies such as artificial intelligence, which present novel and uncertain risks, the absence of established standards and norms of conduct leads to ambiguity regarding the boundaries of liability for negligence. It makes identifying potential risks and determining which behaviour may be deemed wrongful challenging, particularly for users, programmers, and manufacturers. Since these systems are still in develop-

[1441] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 173 f.

[1442] SANDHERR, Strafrechtliche Fragen, 2019, p. 2 f.

[1443] ENGLÄNDER, Das selbstfahrende, 2016, p. 374. For a similar perspective, see: HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 168.

[1444] BECK, Das Dilemma-Problem, 2017, p. 140.

[1445] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 450 Rn. 39.

ment and involve unknown risks, the traditional evaluation of a reasonable person's behaviour[1446] may not provide sufficient guidance either in such complex, technical matters[1447]. The lack of guiding norms further complicates the distinction between permissible and impermissible conduct[1448].

In accordance with the function of negligence in urging individuals to act with greater care and diligence, the actions or omissions necessary to avoid liability for negligent offences can, by their very nature, be uncertain. For instance, the wording of a negligent commission of a crime does not impose a general obligation to act with due care and attention; rather, it establishes a duty to refrain from causing the prohibited outcome. Fulfilling the duty of care represents a means of achieving this objective, while it may not always be sufficient[1449].

Criminal law is not solely concerned with minimising risks; it also enables standardising socially unacceptable behaviours under normative consciousness[1450]. In this regard, the function and role of standards are to serve as significant benchmarks in defining duties of care by balancing the foreseeability of risks with the benefits of a product, accepting residual risks when appropriate, and setting safety requirements to minimise dangers within technical and economic feasibility[1451].

The existence of explicit standards of care is significant in distinguishing between *e.g.* program errors arising from negligent behaviour and those that may occur despite the programmer's best efforts[1452]. In this respect, standards play a crucial role in determining liability, as they establish best practices and formal guidelines to ensure that specific actions align with agreed-upon values. Such standards serve to concretise legal expectations[1453]. Indeed, the uncertainty stemming particularly from negligent liability may cause a chilling effect, deterring firms from developing AI systems, investing in such technologies, engaging in research and development, or even working towards making these systems safer[1454].

[1446] HEGER, StGB § 15 in StGB Kommentar, 2023, Rn. 39.

[1447] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 443 f. Rn. 18.

[1448] BECK, Google Cars, 2017, p. 240, 243.

[1449] JAKOBS, 9. Abschnitt - Strafrecht AT, 1991, p. 319 Rn. 6; GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 577 Rn. 126.

[1450] BECK, Intelligent Agents and Criminal Law, 2016, p. 139.

[1451] VALERIUS, Strafrechtliche Grenzen, 2022, p. 128.

[1452] NISSENBAUM, Accountability in a Computerized Society, 1996, p. 37.

[1453] COOPER, et al., Accountability, 2022, p. 865.

[1454] Singapore, Report on Criminal Liability, 2021, p. 4, [para. 15].

In this regard, the use of flexible general clauses, such as those permitting “development risk” or justifying “socially appropriate use”, could be envisaged as a means to limit the level of care required from the person behind the machine. Alternatively, specific rules and standards could be established to delineate permissible risks across different types and areas of application of AI-driven systems. Such an approach would strike a balance between harnessing the benefits of autonomous systems and ensuring legal certainty, while avoiding unpredictable criminal consequences[1455]. Indeed, the management of risk and uncertainty is not a novel concept in legal discourse, as it can be observed in sectors such as environmental and financial regulation. Establishing foundational principles and liability frameworks to effectively confine risks within acceptable levels serves clarifying duties of care and facilitates the distinction between permissible and impermissible risks[1456].

Undoubtedly, it is crucial to prevent excessive and unjust punishment while ensuring legal certainty for persons behind the machine. It should be possible to determine *ex ante* which risk-creating activities are permissible, and which are impermissible. To achieve this, the required level of care for these systems could be defined for socially beneficial activities, taking into account compliance with the *state of the art*, for instance. By adhering to such established norms of conduct and legal safety standards that define necessary precautions and permissible risks, individuals would be able to gain the orientation and trust needed for conflict-free behaviour without the necessity for additional efforts for hazard prevention[1457]. If all duties of care have been fulfilled, this could be considered within the scope of permissible risk. However, this approach is likely to be criticised both by victims of these crimes and by those who expect technology to be made safer due to the fear of punishment[1458]. Nevertheless, as will be detailed below, it can be argued that it is not feasible to predetermine detailed rules

See also: GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 565.

[1455] HILGENDORF, Gefahr und Risiko, 2020, p. 21; GLESS/WEIGEND, Intelligente Agenten, 2014, p. 591.

[1456] CALO, Robotics and the Lessons, 2015, p. 555.

[1457] ZHAO, Principle of Criminal Imputation, 2024, p. 78 f.
For example, a person driving a car is not required to inspect all of the vehicle’s mechanical components daily; it is sufficient to fulfil what is legally expected from them. See: DUTTGE, Erlaubtes Risiko, 2010, p. 142.

[1458] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 565 f.

for the duty of care in such emerging areas of risk, and this approach risks being reduced to a mere checklist.

The concept of permissible risk itself does not offer a substantive answer on how to define the limits of wrongful actions or who should set these limits[1459]. Legal norms and standards established to define permissible risky behaviour will concretise legal expectations by being incorporated into the duty of care and provide legal certainty. The boundary of the obligation to mitigate risks to a permissible level is open to debate. Indeed, the permissible risk cannot have mathematically precise boundaries; however, it should be as reasonable and transparent as possible. The obligation to mitigate risks cannot be unlimited either, as there is always more that could potentially be done. In this evaluation, a cost-benefit assessment may be taken into account[1460]. However, it should be aligned with the risk-based approach mentioned above[1461] and, in areas such as autonomous driving, it must not be stretched too far when it comes to significant legal interests, such as the life and safety of road users[1462].

Furthermore, these norms should not be subjective but must possess an objective character. They should be determined based on the criteria of foreseeability and preventability, in line with the most advanced scientific knowledge and expertise in the relevant field[1463]. Moreover, they should not only encompass risks and prevention methods that are commonly known but also include those not yet widely recognised, taking into account the knowledge of the few advanced companies operating in the field (in respect of the products manufactured by these companies)[1464]. In this regard, the legal expectations for due care can be concretised, for example, in relation to manufacturers, as adherence to the state of the art, the reasonableness of implementing more stringent protective measures, compliance with technical standards, fulfilment of their own safety assurances (such as those

[1459] MITSCH, Das erlaubte Risiko, 2018, p. 1162.

[1460] ROMANO Leonardo, “Criminal negligence and acceptable risk in the EU’s AI Act: casting light, leaving shadows”, 24.09.2024, https://lawandtech.ie/criminal-negligence-and-acceptable-risk-in-the-eus-ai-act-casting-light-leaving-shadows/.(accessed on 01.08.2025).

[1461] See: Chapter 4, Section C(5)(b)(1)(a)(iii): “Calibrating the Duty of Care Through Risk Levels and Public Tolerance”.

[1462] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 224.

[1463] HOYER, Erlaubtes Risiko, 2009, p. 878.

[1464] TOROSLU/TOROSLU, Ceza Hukuku, 2019, pp. 235-236.

made in advertising), and, finally, meeting the justified expectations of the public[1465].

A licensing procedure, similar to those employed for other activities (such as driving)[1466], could be considered for the development and operation of AI-driven autonomous systems, encompassing all relevant norms of conduct. In light of the risks posed by AI-driven systems, proactive *ex ante* measures should be implemented to prevent harm before it occurs and, accordingly, a licensing system could be applied prior to the commercialisation of such systems, requiring them to meet specific safety and ethical standards[1467]. For instance, licensing for high-risk AI systems might mandate clear requirements related to security, non-discrimination, accuracy, appropriateness, and correctability before they are commercialised[1468]. Furthermore, these licences could be categorised according to the level of risk associated with operating the AI, such as low-risk, high-risk, or systems requiring specialised expertise[1469]. It is argued that systems which are developed using *state of the art* methods and which possess the legally required certification may be assessed under the framework of permissible risk[1470]. Nevertheless, such a certification would merely ensure compliance with certain standards when engaging in risky activities and would not constitute a *carte blanche* for all activities conducted by the licence holder[1471].

Certain partially autonomous systems, such as lane departure warning systems and parking assistance systems, have already been approved by legal systems. Therefore, their use falls within the scope of permissible risk if, *inter alia* the necessary conditions are met. For instance, Section 1a(1) of StVG stipulates that the operation of a motor vehicle using “highly or fully automated driving functions” is permissible if the function is used “as intended”. Section 1a(2) specifies the parameters of its intended use in detail; such as the vehicle being used properly and the driver maintaining control over it in accordance with the specifications[1472]. The manufacturer specifies the conditions under which the system may be used, and the

[1465] HILGENDORF, Moderne Technik, 2015, p. 104.

[1466] THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 284.

[1467] MALGIERI/PASQUALE, Licensing High-Risk AI, 2024, pp. 2-3.

[1468] *Ibid*, p. 2.

[1469] *Ibid*; ASARO, A Body to Kick, 2012, p. 178.

[1470] VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 669.

[1471] See: MAIWALD, Zur Leistungsfähigkeit, 1985, p. 423.

[1472] HILGENDORF, Automatisiertes Fahren und Strafrecht - der Aschaffenburger Fall, 2018, p. 66; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 447 Rn. 31; BECK, Das Dilemma-Problem, 2017, p. 130.

system is required to indicate any usage that deviates from its described parameters[1473].

Through such a regulation, the legislator explicitly addresses permissible risk, ensuring that vehicle manufacturers are not held liable for scenarios that are extremely difficult to recognise[1474]. It is argued that the provisions in the StVG regarding "automated driving" serve as definitions rather than requirements. Vehicles that do not meet these criteria are not legally classified as "highly or fully automated", and, consequently, these rules do not apply to them. In such instances, general traffic laws continue to govern the matter[1475]. Indeed, other standards of due care in road traffic have been further specified, particularly in the German Road Traffic Regulations (StVO) and the German Road Traffic Registration Regulations (StVZO), and referred to in an immense number of court decisions[1476].

(2) Positive Law's Reference to the State of the Science and Technology

Although explicitly established norms and standards aim to define legal expectations and provide clarity, the scope of the duty of care may extend beyond these frameworks. The factors critical for evaluating risks cannot always be fully encompassed by abstract norms. The limit between permissible and prohibited risks can often be ambiguous, and it is impractical for legislators to regulate every detail comprehensively. Assessing permissible risks therefore necessitates looking beyond the mere text of the law

[1473] GREGER, Haftungsfragen, 2018, p. 2.

[1474] STEINERT, Automatisiertes Fahren, 2019, p. 6.

[1475] HILGENDORF, Automatisiertes Fahren und Strafrecht - der Aschaffenburger Fall, 2018, p. 66.

[1476] HEGER, StGB § 15 in StGB Kommentar, 2023, Rn. 39b.
In Turkish law, certain regulations concerning autonomous vehicles were introduced through a by-law, prepared in alignment with European Union legislation (Commission Implementing Regulation (EU) 2022/1426 of 5 August 2022). See: "Tam Otonom Araçların Otonom Sürüş Sistemine İlişkin Motorlu Araçların Tip Onayı Hakkında Yönetmelik", Official Journal on 01.12.2024 (Issue No. 32739), https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=41078&MevzuatTur=7&MevzuatTertip=5). See also: "Motorlu Araçlar ve Römorkları İle Bunlar İçin Tasarlanan Aksam, Sistem ve Ayrı Teknik Ünitelerin Genel Güvenliği Ve Korunmasız Karayolu Kullanıcılarının ve Yolcuların Korunması İle İlgili Tip Onayı Yönetmeliği", Official Journal on 14.05.2020 (Issue No. 31127), https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=34512&MevzuatTur=7&MevzuatTertip=5. (accessed on 01.08.2025).

to other overarching legal principles[1477]. As such, legal safety standards are frequently not exhaustive and may require further interpretation or clarification[1478]. These standards or guidelines often have a generalising effect, which may prove inadequate in specific cases where more tailored conduct is necessary. Additionally, they may become outdated over time[1479]. Similarly, in the context of sports competitions, not all potential actions can be meticulously regulated. As a result, the scope of unregulated actions is often considerable, which leaves room for interpretation and adaptation to the particular circumstances[1480].

Given the impracticality of regulating every individual scenario within the scope of risk management, legislators often utilise concepts such as the "state of the science or technology"[1481], or delegate risk assessment to the executive body. By incorporating such provisions, they establish a framework for both the approval of hazardous activities and the determination of the obligations of persons behind the machine[1482]. The reference to the current state of science and technology considers the rapidly evolving development of emerging technologies, such as AI-driven autonomous systems, and ensures that legally standardised due care obligations keep up with the pace of this progress, preventing them from becoming outdated quickly[1483].

Indeed, listing specific standards for each application or referencing "generally recognised rules of technology" may cause the legal system and the measures to be implemented to lag behind the latest advancements in science and technology. This is because technology evolves at an exceptionally rapid pace, which makes static references insufficient to address emerging developments effectively[1484]. With every technical innovation, new technical norms of conduct are formulated in advance of an actual

[1477] MITSCH, Das erlaubte Risiko, 2018, p. 1165.
The provisions concerning negligent liability (e.g., Section 222 of the StGB) are general and open-ended, encompassing the technical norms of safety-related conduct. However, where more specific standards exist, they will apply in determining the scope of negligence, in accordance with the principle of the precedence of more specific norms. See: IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 295 f.

[1478] FRISTER, 10. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 129 Rn. 11.

[1479] FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 182 Rn. 57.

[1480] GIEZEK, Einige Bemerkungen, 2009, p. 547.

[1481] CORNELIUS, Künstliche Intelligenz, 2020, p. 59.

[1482] SCHÖMIG, Gefahren und Risiken, 2023, p. 201.

[1483] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 227.

[1484] HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 227.

violation of a legal interest[1485]. Furthermore, such explicit and detailed rules may conflict with the abstract and general structure of the criminal code, result in overly complex and confusing regulations that fail to clearly indicate criminal liability, require constant updates due to technological advancements, and hinder innovation through lengthy adjustment procedures[1486]. Therefore, by employing concepts such as the "state of the science or technology", the perspective of an expert possessing the most up-to-date technical or scientific knowledge is taken into account[1487]. The greater the control over the risks, the stricter the rules for due care become[1488]. In this context, the standard of the duty of care is adjusted to align with the evolving risk threshold, meaning behaviour considered cautious today may no longer meet that standard if the risk threshold changes[1489]. For example, in the case of products, the time when the manufacturer places the product on the market is taken into account[1490].

In some cases, legislation explicitly refers to generally recognised rules of technology or the state of the science or technology when determining the scope of the duty of care. For instance, pursuant to Section 5(1)(2) of the *Bundesimmissionsschutzgesetz* (BImSchG)[1491], installations subject to licensing are required to be constructed in accordance with the *state of the technique*. Similarly, according to Section 16(1) of the *Gentechnikgesetz* (GenTG)[1492]; *"(1) Approval for a release must be granted if 1. the requirements in accordance with (...) are met, 2. it is guaranteed that all safety*

[1485] IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 145.

[1486] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 262.

[1487] HOYER, Erlaubtes Risiko, 2009, p. 872.

[1488] HILGENDORF, Digitalisierung, Virtualisierung und das Recht, 2020, p. 409.

[1489] GIEZEK, Einige Bemerkungen, 2009, p. 549.
If the objective standard of state of the technology were to be applied in the context of Sections 222 and 229 of StGB to products that do not require approval, it would still necessitate that the objective dangerousness of a particular technology was at least subjectively recognisable to the perpetrator. See: HOYER, Erlaubtes Risiko, 2009, p. 877.

[1490] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 9.

[1491] Gesetz zum Schutz vor schädlichen Umwelteinwirkungen durch Luftverunreinigungen, Geräusche, Erschütterungen und ähnliche Vorgänge (Bundes-Immissionsschutzgesetz - BImSchG), enacted on 15.03.1974, last amended on 03.07.2024,https://www.gesetze-im-internet.de/bimschg/BJNR007210974.html. (accessed on 01.08.2025).

[1492] Gesetz zur Regelung der Gentechnik (Gentechnikgesetz - GenTG), enacted on 20.06.1990, last amended on 27.09.2021, https://www.gesetze-im-internet.de/gentg/BJNR110800990.html. (accessed on 01.08.2025).

precautions required according to the state of science and technology are taken, 3. According to the state of science, unacceptable harmful effects on the legal interests specified in Section 1 No. 1 are not to be expected in relation to the purpose of the release". Additionally, Section 9(2)(3) of *Atomgesetz* (AtomG)[1493] requires that "*the necessary precautions against damage caused by the use of nuclear fuel have been taken in accordance with the state of science and technology*" as a condition for obtaining a license, among other requirements[1494].

Section 3(6) of the *Bundesimmissionsschutzgesetz* (BImSchG) defines state of the technology as: *"(…) the state of development of advanced processes, equipment or operating methods which appears to ensure the practical suitability of a measure for (…) or otherwise for avoiding or reducing impacts on the environment in order to achieve a generally high level of protection for the environment as a whole"*[1495]. In this regard, the distinction between the state of the science and the state of the technology lies in their respective approaches to risk management. The state of the technology mandates the use of technically feasible methods to minimise risks. If no alternative course of action with a lower risk is currently known, it is presumed that the necessary precautions have been taken. In contrast, the state of the science considers whether any technological solution exists to sufficiently mitigate the risks of a particular action. If no such technology is available, the action may be deemed excessively risky relative to its anticipated social benefits and would therefore not be authorised[1496].

Finally, the question of who should draft the content of standards is of critical importance. This issue becomes particularly significant when generally recognised rules of technology are to be established as standards. While private parties may also draft such rules, this could raise other concerns. The lawmaker can refer to the content of a specific set of these technical rules and, in a sense, incorporate them into legal norms. Nonetheless, it must be recognised that this approach could lead to challenges arising from regulating a static set of rules that lack the required dynamism to adapt to technological advancements and it would inherently fail to align

[1493] Gesetz über die friedliche Verwendung der Kernenergie und den Schutz gegen ihre Gefahren (Atomgesetz), enacted on 23.12.1959, last amended on 04.12.2022, https://www.gesetze-im-internet.de/atg/BJNR008140959.html. (accessed on 01.08.2025).

[1494] HOYER, Erlaubtes Risiko, 2009, p. 865 ff.

[1495] Translation has been made by the author.

[1496] HOYER, Erlaubtes Risiko, 2009, p. 865, 873.

with the rapid advancements in technology. Thereby it makes the state incapable of performing its constitutional obligation to protect the welfare of its citizens[1497].

(3) The Effectiveness of Norms Established by Private Entities on the Duty of Care

The necessity for numerous diverse norms of conduct, along with their continuous evolution, makes it impractical for the state to regulate and consistently update standards and safety guidelines applicable in every field[1498]. Furthermore, due to its remoteness from specific fields, the state may be unable to establish ideal instructions on such matters. Therefore, not all norms of conduct are established by official authorities[1499]. Private entities, such as professional associations, federations, and civil organisations, frequently develop detailed rules that function as standards within their respective fields. Compliance with these standards -whether written or unwritten- can influence legal assessments of duty of care[1500]. Such non-governmental industry standards play a significant role, and official regulations occasionally refer to them. While reflecting the current state of science and technology, they do not establish new benchmarks but merely report the existing situation[1501].

One of the best examples of certain social groups establishing their own rules with government approval (self-regulation) is found in sports competitions. Although the legislator does not prescribe any rules for the practice of sports and leaves it to the autonomy of the sports associations, it is not a criminal law-free area[1502]. However, a significant difference between sports competitions and other risky activities, such as road-traffic, lies in the fact that traffic rules are more explicitly and comprehensively regulated[1503]. Besides, the risks associated with sports competitions generally concern

[1497] *Ibid*, p. 869 f.
[1498] LENCKNER, Technische Normen, 1969, p. 490.
[1499] TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 235.
[1500] EISELE, §12 Die Fahrlässigkeit, 2016, p. 303 Rn. 32.
[1501] VALERIUS, Sorgfaltspflichten, 2017, p. 10 f.
[1502] MITSCH, Das erlaubte Risiko, 2018, p. 1165.
[1503] HEGER, StGB § 15 in StGB Kommentar, 2023, p. 51.

only those directly involved, whereas fields such as automotive, industry, and AI pose risks that extend to uninvolved individuals as well[1504].

In Germany, for instance, the standards issued by bodies such as the *Deutsches Institut für Normung* (DIN), *Verband der Elektrotechnik, Elektronik und Informationstechnik* (VDE), De*utscher Verein des Gas- und Wasserfaches* (DVGW), and *Verein Deutscher Ingenieure* (VDI) guide the production and application of technologies. Developed by private associations, they ensure safety, simplify processes, and address risks associated with advancing technologies, while promoting industrial progress[1505]. Similarly, in Turkey, the Turkish Standards Institution (*Türk Standartları Enstitüsü* - TSE)[1506] and, globally, the International Organization for Standardization (ISO)[1507] play significant roles in the development and establishment of standards. To illustrate, the "ISO/IEC 42001:2023 Standard", provides a comprehensive framework for establishing, implementing, maintaining, and continually improving AI management systems; and addresses key issues such as ethical considerations, transparency, accountability, and risk management to ensure the responsible and trustworthy use of AI technologies[1508]. There may be alignment issues between standards established by different organisations at varying levels. For instance, national standards may be either softer or stricter compared to EU standards[1509]. It can be argued that, in such cases, the stricter and more comprehensive standards should be applied to mitigate risks; as otherwise, it would constitute a violation of the stricter standards.

Undoubtedly, in the performance of certain tasks, both written and unwritten rules, such as established professional norms, are as important as formal guidelines and standards, as they demonstrate the optimum behavioural expectations for due care. However, particular attention must be paid to this issue in the context of high-risk technologies with the potential to fundamentally alter societal dynamics, such as AI-driven autonomous systems. This is because the actors involved in the formation of standards

[1504] For discussions on the evaluation of typical and atypical risks concerning permissible risk in the context of sports competitions, see: Chapter 4, Section C(5)(b)(1)(b): "The Relationship Between Social Adequacy and Permissible Risk".

[1505] LENCKNER, Technische Normen, 1969, p. 490.

[1506] https://www.tse.org.tr. (accessed on 01.08.2025).

[1507] https://www.iso.org. (accessed on 01.08.2025).

[1508] ISO/IEC 42001:2023 Information Technology - Artificial intelligence - Management system, 1st edition., 2023, https://www.iso.org/standard/81230.html. (accessed on 01.08.2025).

[1509] LENCKNER, Technische Normen, 1969, p. 492.

may not only aim to mitigate risks to legal interests but also act to protect their own economic and other interests[1510]. Moreover, standards must be set at a high level, as AI-driven systems may pose extraordinary risks to social life[1511].

The extent to which standards established by non-state entities should be considered in determining the duty of care is a subject of debate. Some views assert that non-state industry standards cannot serve as a source for determining the duty of care, and relying on private standards to determine negligence is inconsistent, as these norms are created by non-authoritative bodies and may not hold clear legal or evidentiary weight. On the other hand, the counter-argument asserts that well-established norms reflect practical, proven practices that indicate appropriate care without solely determining it, making them valuable guides in assessing negligence[1512]. Thus, the industry standards, self-commitment of the responsible person, general social ethics, professional ethics, and similar factors can indicate the scope of the duty of care[1513].

Indeed, non-state rules from the respective social context, such as ISO or DIN standards, reflect the required care to be exercised in certain activities and, in this regard, serve as an important indicator for determining the duty of care[1514]. However, such technical standards do not have a binding effect on courts. Behaviour contravening these rules cannot be directly equated with a failure to exercise due care. Individuals subject to such norms must critically assess whether the standards adequately address the specific risks involved, as these norms may have become outdated and fail to incorporate the latest advancements in the field. Consequently, the standard of care required might exceed the guidelines set by the existing technical criteria[1515]. In this regard, technical descriptions should not be confused with legal standards of care, which are determined by legislators and courts[1516].

[1510] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 444 Rn. 20.

[1511] KAIAFA-GBANDI, Artificial intelligence, 2020, pp. 315 – 316.

[1512] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1163, Rn. 223.

[1513] ZHAO, Principle of Criminal Imputation, 2024, p. 87.

[1514] KASPAR, Grundprobleme, 2012, p. 20; BECK, Das Dilemma-Problem, 2017, p. 123 f.

[1515] HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 449.

[1516] HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 153.

It is imperative that criminal law considers the collective legal interests of society and does not merely enforce the stipulations of non-state entities. Industry standards and safety guidelines, while valuable as guidance and in civil contexts, are not legally binding in criminal assessments and are typically designed with civil liability in mind[1517]. While these norms can serve as indicators of whether an individual's behaviour aligns with the legal standard of care, they are rebuttable and may be insufficient to fully address the specific circumstances of a given case. Thus, violations of specific non-criminal provisions, such as safety regulations, may suggest a lack of due care but require careful consideration within the distinct framework of criminal law[1518].

(4) Compliance with Norms: An Indicator of Fulfilling the Duty of Care

The concept of duty of care, central to the analysis of liability arising from negligence, may stem from a wide variety of sources[1519]. The determination of whether an individual has breached their duty of care often involves consideration of numerous and, in some cases, unwritten sources[1520]. Among these, alongside statutory regulations, there may be safety measures designed to mitigate the risks associated with specific hazardous activities, as well as non-legal norms such as technical standards, requirements stemming from the inherently dangerous nature of certain activities, or generally recognised principles of experience[1521]. The reliance on a range of such norms creates significant uncertainty, which in turn undermines an individual's ability to regulate their behaviour accordingly. Besides, in such an uncertain environment, the potential for criminal sanctions causes

[1517] BECK, Intelligent Agents and Criminal Law, 2016, p. 139.

[1518] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 183 Rn. 51; BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 444 Rn. 21; VELLINGA, Cyber Security, 2023, p. 135.

[1519] This issue is examined in detail above. See: Chapter 4, Section C(4): "The Scope and Boundaries of Duty of Care for the Person Behind the Machine".

[1520] VALERIUS, Sorgfaltspflichten, 2017, p. 21.

[1521] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1143, Rn. 172 f.; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1125; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 531 Rn. 16 f.; STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 310 Rn. 19 f.
For an analysis of these in determining the permissible risk, see: ÜNVER, Ceza Hukukunda İzin Verilen Risk, 1998, p. 364.

a significant deterrence for manufacturers and developers of AI-driven systems[1522].

To mitigate uncertainty, it may be considered necessary to develop clear and precise criteria to delimit the scope of criminally relevant duties of care. However, identifying such criteria presents significant challenges. One potential, though, far-reaching approach would be to restrict criminal liability to duties of care explicitly defined by law; because even a general reference to the "state of science and technology" would be overly vague. Alternatively, criminal liability could be confined to breaches of essential duties of care. While the term "essential" itself remains imprecise, it would nonetheless serve as an initial constraint on what might otherwise be excessively broad duties of care[1523].

In cases where the duty of care is explicitly defined by special norms, the question arises as to whether the persons behind the machine can exculpate themselves by demonstrating compliance with the relevant technical standards, or conversely, whether negligence can be established solely on the grounds that they failed to meet those technical standards[1524]. Indeed, in practice, many researchers and manufacturers operate under the belief that they are acting lawfully by adhering to established standards[1525]. However, is this truly the case? According to one view, if all such norms of conduct and specific measures intended to prevent the harmful outcome are explicitly enumerated, and if the individual fully complies with the measures defined in these norms, no liability arises. However, if the norm does not enumerate all preventive measures explicitly, merely listing some of them as examples or imposing a general duty to take precautionary measures, compliance with these alone does not absolve the individual of liability[1526].

[1522] BECK, Das Dilemma-Problem, 2017, p. 129.
One perspective in the debate on whether reliance on unwritten norms in determining the duty of care violates the principle of legal certainty asserts that this is not the case. According to this view, as long as the conditions of care are not overly expanded and their content is concretely supported by additional legal norms, this approach is more appropriate -particularly in technical matters where scientific progress is rapid -and does not contravene the constitution. See: DEMIREL, Taksir, 2024, p. 772.

[1523] For the discussion, see: VALERIUS, Sorgfaltspflichten, 2017, p. 21.

[1524] LENCKNER, Technische Normen, 1969, p. 491 f.

[1525] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 444 Rn. 22.

[1526] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 135 f.; SCHÖMIG, Gefahren und Risiken, 2023, p. 149 ff.; ZAFER, Ceza Hukuku, 2021, p. 351; ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 285.

Approaching the issue from a different perspective, it can be argued that if standards of safety precautions or care have been established in a particular area, this supports the assertion that a legally relevant risk exists[1527]. Specific rules, such as those set out in regulations like the StVO, establish conditions for certain risky activities. Adherence to these rules generally indicates that an individual is not creating a legally disapproved danger. The breach of such technical standards, professional rules, and other informal regulatory systems indicates the creation of an impermissible risk[1528]. In this regard, it can be argued that compliance with these rules principally precludes any objectively negligent dangerous behaviour at the initial level (i.e., the primary assessment of wrongfulness) and the corresponding criminal liability, as the legislator has explicitly excluded the consideration of such risks. However, when an additional factor comes into play, the individual may need to exercise even greater caution in light of this circumstance. For instance, if there is an obstacle on the road, merely adhering to the 30 km/h speed limit would not suffice; the driver must reduce their speed further[1529].

Although it does not directly pertain to criminal law, the German Product Liability Act (*Produkthaftungsgesetz* - ProdHaftG)(Section 1(2)(4) and (5) provides that the manufacturer shall not be held liable if the defect arose because the product complied with mandatory regulations at the time it was placed on the market, or if the defect could not have been detected based on the state of science and technology at the time the product was

For example, under Turkish law, according to a provision in the Construction Zoning Law (*İmar Kanunu*), Article 28(11), if the owner of a building under construction does not assume any roles (such as construction contractor, or site supervisor for a structure with a valid permit) all liability rests, as appropriate, with the project owners, the construction contractor, the site supervisor, and other relevant technical personnel. Based on this regulation, it is argued that if the construction of a building is carried out under the responsibility, supervision, and control of an officially certified engineer with the necessary expertise, then they are held liable for any crimes resulting from a technical collapse of the building. However, in accordance with this regulation, the building owner is not held liable, as they are deemed to have fulfilled their duty by entrusting the task to a duly qualified professional. See: ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 278.
For the provision, see: İmar Kanunu (Nr. 3194), Official Journal on 09.05.1985 (Issue No. 18749), https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=3194&MevzuatTur=1&MevzuatTertip=5. (accessed on 01.08.2025).

[1527] ROXIN/GRECO, § 11. Die Zurechnung in Strafrecht AT, 2020, p. 489 Rn. 67.

[1528] JAKOBS, 7. Abschnitt - Strafrecht AT, 1991, p. 205 Rn. 44.

[1529] KAIAFA-GBANDI, Artificial intelligence, 2020, p. 320.

introduced into circulation. Similarly, Article 11(1)(d) and (e) of the new EU Product Liability Directive (PLD) contains comparable provisions, stipulating that the manufacturer shall not be held liable if the defect that caused the damage was due to the product's compliance with "legal requirements"[1530]. In this regard, one perspective argues that the manufacturer should be able to exonerate themselves if the vehicle has been approved in accordance with the legally relevant state of science and technology and if the manufacturer does not possess superior expert knowledge[1531].

Despite these discussions, it is important to recall the key features of criminal law. The negative formulation of norms of conduct does not imply a positive assumption that anything not explicitly prohibited is permitted. This is because the relevant regulations may be incomplete or, as in Section 1(2) of the StVO[1532], include a general prohibition against causing harm[1533]. To illustrate, in a location with a speed limit of 90 km/h, a driver traveling at 80 km/h encounters a pedestrian who suddenly jumps into the road, resulting in a fatal collision. In this context, compliance with the 90 km/h speed limit does not amount to a general permit allowing the driver to act without further consideration. If the driver adheres to all specific norms and observes the general principle of refraining from causing harm, and the accident remains unavoidable, only then does the concept of permissible risk apply[1534]. Thus, in accordance with Section 1 of the StVO, in a specific situation where it is evident, foreseeable and avoidable that harm will result, the person causing the harm cannot escape liability by merely claiming compliance with the rules[1535].

In this regard, permissible risk does not grant the actor a *carte blanche*. Even when acting within the generally permissible limits, this does not absolve them from the obligation to take additional precautions in specific situations beyond what general standards of care require. If the realisation of the risk is foreseeable in a particular circumstance, the actor has a duty to prevent it, provided they are still in a position to avert the harmful

[1530] For an evaluation, see: VELLINGA, Cyber Security, 2023, p. 135.

[1531] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 224.

[1532] Translation is made by the author: "Whoever participates in the road traffic must behave in such a way that no other person is harmed, endangered or more than unavoidably inconvenienced or harassed under the circumstances."

[1533] JAKOBS, 7. Abschnitt - Strafrecht AT, 1991, p. 205 Rn. 45.

[1534] For a different evaluation, see: KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 404.

[1535] MAIWALD, Zur Leistungsfähigkeit, 1985, p. 421.

outcome at that moment[1536]. Indeed, legally defined standards of duty of care (*normierte Sorgfaltspflichten*) serve as a baseline, but they are not absolute. They can be exceeded depending on the specific circumstances and potential risks involved. Fulfilling the duty of care may require a wide range of possible actions[1537].

In negligence-based liability, whether due care has been exercised should be assessed based on the specific circumstances of each individual case, rather than relying exclusively on abstract rules[1538]. In certain situations, it may even be necessary to act contrary to general guidelines or rules if the specific context so requires[1539]. For instance, if children are playing on the right side of the road, it may be necessary to drive on the left, even if this deviates from the relevant rule[1540]. Similarly, compliance with the norm does not always suffice. For instance, the driver mentioned above travelling at 80 km/h on a road with a 90 km/h speed limit must reduce their speed if faced with a potential accident risk. Failing to do so (even if such a general duty is not explicitly stipulated in road traffic legislation) breaches the duty of care, potentially leading to negligence-based liability[1541]. Observance of the objective duty of care cannot be made a reason for excluding wrongdoing by itself[1542], and rule-compliant behaviour does not exempt one from adhering to the prohibition of harming others[1543]. This is because, in addition to specific rules, the general principle of not causing harm to others prevails and the incident must be evaluated with all its details[1544].

The general principle of refraining from causing harm, while explicitly enshrined in general prohibitions such as Section 1 of the StVO, is also applicable beyond road traffic. Indeed, even when specific standards are

[1536] *Ibid*, p. 423.

[1537] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1143, Rn. 172 f.

[1538] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 311 Rn. 21.

[1539] VALERIUS, Sorgfaltspflichten, 2017, p. 11.

[1540] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1144, Rn. 174.

[1541] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1123; DEMIREL, Taksir, 2024, p. 85.

[1542] OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 246.

[1543] DUTTGE, Erlaubtes Risiko, 2010, p. 145.

[1544] HORN, Erlaubtes Risiko, 1974, p. 725; MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 176.
See also: ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1196 Rn. 36.

followed, exceptional cases may still reveal a lack of due care or impermissible risky behaviour, as even the most detailed harm-mitigation regulations may prove insufficient. Particularly in cases involving biased, outdated, or otherwise inapplicable regulations, adherence to provisions based on the legislature's apparent misjudgement may lead to harmful outcomes[1545]. Such instances serve as notable examples. Exceptions, however, are conceivable where, despite a breach of the regulation, adequate alternative safety measures are implemented, or where the breached regulation addresses risks other than those that actually materialised[1546]. In such cases, it must be examined whether the incident falls within the protective scope of the norm. If it does, liability for negligence may arise[1547].

In conclusion, it is essential to emphasise that the aforementioned norms of conduct and special rules play a crucial role in determining the requisite standard of care and reducing risks in the performance of tasks. A breach of duty generally arises when the perpetrator fails to adhere to the prescribed legal standards of behaviour, unless the circumstances deviate from what the norm intended, or the norm itself has become outdated. However, compliance with standards of care serves merely as an indicator of the absence of negligence and does not conclusively establish it[1548]. Similarly, compliance with such rules does not necessarily absolve an individual of liability[1549]. In non-regulated areas of life, the same function is fulfilled by the model of a prudent and conscientious person in the same situation and social role[1550].

In other words, compliance with such norms merely constitutes an indicator that the duty of care has been fulfilled. Negligence may still be established even if these rules are followed[1551]. Beyond this, in all cases, it

[1545] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1189 Rn. 18 ff.

[1546] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1162 f., Rn. 222.

[1547] HARDTUNG, StGB § 222 MüKo, 2021, Rn. 19.

[1548] STERNBEG-LIEBEN/SCHUSTER, StGB § 15 Vorsätzliches und fahrlässiges Handeln in Schönke/Schröder Strafgesetzbuch, 2019, Rn. 135 f.; SCHÖMIG, Gefahren und Risiken, 2023, p. 149 ff.

[1549] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1123; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 531 Rn. 16 f.

[1550] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 310 Rn. 19 f.; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1125; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 223 Rn. 20.

[1551] WELZEL, Das deutsche Strafrecht, 1969, p. 131 f.; EISELE, §12 Die Fahrlässigkeit, 2016, p. 304 Rn. 35; HILGENDORF, Moderne Technik, 2015, p. 110 fn. 43; HARD-is essential to examine whether other norms falling within the scope of the duty of care are applicable in light of the specific circumstances of the case, and most importantly, whether the general principle to refrain from harm has been upheld. Thus, the behavioural rules are supplemented, or even overridden, by the principle of best possible avoidance of harm to legal interests[1552]. Particularly, exceptional circumstances that significantly heighten the risk in a given situation may give rise to duties of care that go beyond the usual standard[1553].

Risk management systems that operate by following standards and established norms are highly important; however, they may fail to prevent harmful outcomes by creating an illusion of acceptable risks and reducing the pursuit of trustworthy AI to mere compliance via "box-ticking" rather than substantive safety[1554]. Therefore, while such violations can indicate negligence, courts must independently assess the actual risk created, and compliance with these norms does not necessarily preclude the existence of disapproved danger, especially in exceptional cases that demand stricter standards[1555]. The determination of the appropriate duty of care in individual cases primarily falls within the sphere of legal practice and is assessed on a case-by-case basis[1556].

(5) The EU AI Regulation (AI Act) and the Imposed Duty of Care

The inherently cross-border nature of digitalisation and AI, due to its nature and scope, necessitates establishing international or supranational regulations to ensure effective governance and responsibility[1557]. The EU

TUNG, StGB § 222 MüKo, 2021, Rn. 18; SCHÖMIG, Gefahren und Risiken, 2023, p. 150.
This view is also recognised in Turkish law. See: DEMIREL, Taksir, 2024, p. 85.

[1552] EISELE, §12 Die Fahrlässigkeit, 2016, p. 303 Rn. 33.

[1553] HARDTUNG, StGB § 222 MüKo, 2021, Rn. 20; KASPAR, Grundprobleme, 2012, p. 20; SCHÖMIG, Gefahren und Risiken, 2023, p. 149 ff.
See also: DUTTGE, StGB § 15 MüKo, 2024, Rn. 104.

[1554] ROMANO Leonardo, "Criminal negligence and acceptable risk in the EU's AI Act: casting light, leaving shadows", 24.09.2024, https://lawandtech.ie/criminal-negligence-and-acceptable-risk-in-the-eus-ai-act-casting-light-leaving-shadows/.(accessed on 01.08.2025).

[1555] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1189 Rn. 18 ff.

[1556] SCHÜNEMANN, Moderne Tendenzen, 1975, p. 578; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 262 f.

[1557] ROBLES CARRILLO, Artificial Intelligence, 2020, p. 15.

AI Regulation[1558], commonly referred to as the *AI Act*, represents the most comprehensive legal framework on artificial intelligence to date. Whether this regulation, as observed in the EU's General Data Protection Regulation (GDPR), will set a global benchmark for AI governance and risk-based approach through the phenomenon known as *Brussels Effect* remains to be seen[1559].

With respect to criminal liability, neither the AI Regulation nor the AI Liability Directive (AILD), as previously mentioned[1560], offers any explicit guidance. Indeed, it would be unreasonable to expect such supranational legal text, particularly in the form of a Regulation, to address this issue. Nevertheless, the AI Regulation imposes certain restrictions on the production, utilisation and deployment of certain AI systems. In this regard, this section will examine whether it provides any guidance in determining the duty of care concerning criminal liability in offences involving AI-driven systems. In other words, it should be examined whether the provisions of the AI Regulation could be considered in assessing whether the duty of care has been breached in cases where a high-risk or limited-risk AI-driven system causes injury to an individual.

The AI Regulation adopts a risk-based approach, categorising AI applications into different risk classes. Risk-based approaches ensure that duties and obligations are aligned with the level of actual risk by prioritising and calibrating enforcement actions in a manner that is proportional to the nature of the specific hazards[1561]. Indeed, the risk-based approach is not a novel concept. In the EU, particularly since the introduction of the Digital Single Market Strategy, various risk-based approaches have been consistently employed to regulate the digital economy, notably in areas such as data, online content, platforms, cybersecurity, digital products and services, and AI[1562].

[1558] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Regulation), 12.07.2024, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689. (accessed on 01.08.2025).

[1559] GRAHAM/THANGAVEL/MARTIN, Navigating AI-Lien Terrain, 2024, p. 203.

[1560] See: Chapter 3, Section C(1)(c)(4): "The EU AI Liability Directive (AILD) and Strict Liability Regime within the EU".

[1561] EBERS, Truly Risk-Based, 2024, p. 4.

[1562] *Ibid*, p. 4 f.

Nevertheless, the AI Regulation does not follow a truly risk-based approach due to, *inter alia*, the absence of a risk-benefit analysis, limited reliance on empirical evidence and abstract risk-categories[1563]. The framework largely overlooks the benefits and positive contributions of AI systems, focusing primarily on risk prevention[1564]. As a result, it neither incorporates a risk-benefit analysis nor clearly addresses whether a certain level of risk can be deemed acceptable in light of the societal gains offered by AI (-driven) systems[1565]. However, since no one wishes to be harmed unnecessarily, society accepts certain risks in pursuit of potential benefits; therefore, a risk-based approach should consider both negative and positive effects[1566]. The risk categories adopted in the Regulation are pre-defined. As a result, certain applications are classified as high-risk AI systems under *Annex III* merely because of their use in specific sectors and purposes, even if they do not pose a significant risk of harm, while some of the most dangerous systems, such as military killer robots, remain outside its scope[1567].

The current regulatory approach is market-driven. The primary objective of the (proposed) AI regulatory frameworks within the EU (the AI Regulation and the AI Liability Directive)[1568] is to facilitate the unrestricted commerce of AI technologies while addressing extreme risks[1569]. Rather than pursuing another approach to eliminate all risks or reduce risks to an acceptable level, the frameworks adopt a proportionate regulatory approach. This aims to strike an optimal balance between two key objectives: mitigating the risks associated with AI (-driven) systems and fostering innovation to maximise their benefits. By seeking to minimise potential harms while accounting for the costs of regulation, the approach

[1563] *Ibid*, p. 11.

[1564] For a different risk-based approach, see: SCHÖMIG, Gefahren und Risiken, 2023, p. 270 ff.
For the risk-based approach adopted in this study, see: Chapter 4, Section C(5)(b)(1): “Risk-Based Approach”.

[1565] EBERS, Truly Risk-Based, 2024, p. 12 f.

[1566] *Ibid*, p. 9.

[1567] *Ibid*, p. 15.

[1568] See also: European Parliament. Resolution of 16 February 2017 on Civil Law Rules on Robotics (2015/2103(INL)), Official Journal of the European Union, https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf. (accessed on 01.08.2025).

[1569] RESTREPO AMARILES/BAQUERO, Promises and Limits of Law, 2023, p. 6.

ensures that safety measures do not unnecessarily impede technological progress[1570].

The main advantages of establishing risk classes in risk-based approaches, lie in their ability to systematise complex decision-making processes, ensuring evaluations are both predictable and adaptable to individual cases. However, the disadvantages include criticisms of being overly vague, excessively complex, or prone to subjective interpretations, which may, in turn, hinder innovation in emerging technologies[1571]. Nevertheless, the EU Regulation partially addresses this issue, particularly for high-risk AI systems, by providing an exhaustive list. Yet, this approach is still criticised as impractical due to the complexity and evolving nature of AI technology, which makes strict classification challenging. Additionally, the risk of information asymmetry between developers and regulators may further hinder accurate risk assessment[1572].

The AI Regulation employs a four-tiered classification for AI systems, based on the level of risk they present. These are: "unacceptable", "high", "limited" and "minimal" risk. While minimal-risk AI systems, including the majority of standard AI applications, are subject to few or no additional requirements; limited-risk AI systems, such as chatbots, are required to implement transparency measures to ensure that users are aware that they are interacting with a machine. The high-risk AI category includes applications in essential areas like medical diagnostics, critical infrastructure, education or employment. These systems are subject to strict obligations and requirements concerning transparency, data governance, and human oversight. Finally, the category of unacceptable-risk AI encompasses systems that can manipulate human behaviour or exploit vulnerable groups, which are explicitly prohibited.

In the context of the AI Regulation, the central debate concerns whether the obligations and requirements imposed on high-risk and limited-risk systems can serve as a source of the duty of care, the breach of which could give rise to liability for negligence under national law. Indeed, the AI Regulation, particularly Section 2, under Article 8 and the following provisions, imposes various requirements for high-risk AI systems to providers, such as

[1570] EBERS, Truly Risk-Based, 2024, p. 9.

[1571] SCHÖMIG, Gefahren und Risiken, 2023, p. 285 f.

[1572] HEISS, Künstliche Intelligenz, 2021, p. 2; SCHÖMIG, Gefahren und Risiken, 2023, p. 276.

establishing a risk management system (Art. 9)[1573], ensuring human oversight (Art. 14), and providing instructions for use (Art. 13(2)). Additionally, data governance must be implemented to ensure that training, validation, and testing datasets are relevant, adequately representative and, as far as possible, error-free (Art. 10(3)). Consequently, the implementation of these measures serves to mitigate the risks associated with the utilisation of AI (-driven) systems, by reducing both the probability of adverse events occurring and the potential severity of any such occurrences.

Additionally, certain obligations are also imposed on other actors, such as deployers. For instance, they are required to take appropriate technical and organisational measures to ensure that the systems are used in accordance with the provided instructions, and to assign human oversight to natural persons with the necessary competence, training, and authority, as stipulated in Article 26. Furthermore, certain obligations are imposed on providers of "general-purpose AI [(GPAI)] models with systemic risk" under Section 3, Article 55. Accordingly, providers of such GPAI models must conduct model evaluations, including adversarial testing, to identify and mitigate risks; assess and address potential systemic risks and their sources; promptly track, document, and report serious incidents and corrective measures to the AI Office and relevant authorities without undue delay; and ensure an adequate level of cybersecurity protection[1574].

Since each of these obligations and requirements would require separate academic analysis, they will not be discussed in detail here to avoid exceeding the scope of this study. What is essential to emphasise, however, is that the AI Regulation seeks to mitigate the risks posed by AI (-driven) systems through these obligations and requirements. Therefore, implementing and complying with these provisions can be considered as part of the duty of care owed by persons behind the machine. In other words, a failure by the actors addressed under the AI Regulation to fulfil these obligations and requirements may constitute a breach of the duty of care, potentially giving rise to liability for negligence.

[1573] It is argued that this provision aims to ensure that, through appropriate and targeted risk management systems, providers of high-risk AI systems reduce risks to a residual level after all precautions have been taken, thereby making the remaining risk permissible. See: ROMANO Leonardo, "Criminal negligence and acceptable risk in the EU's AI Act: casting light, leaving shadows", 24.09.2024, https://lawandtech.ie/criminal-negligence-and-acceptable-risk-in-the-eus-ai-act-casting-light-leaving-shadows/.(accessed on 01.08.2025).

[1574] For the full text of the provision, see Article 55 of the AI Regulation.

Nevertheless, not all obligations and requirements imposed on these actors can be regarded as part of the duty of care in relation to a specific criminal offence. For instance, the logging and record-keeping requirement outlined in Article 12 has no direct relevance to preventing harmful outcomes in a specific incident, as it primarily serves to assist in illuminating the event *ex post*. Similarly, the technical documentation requirement under Article 11 does not directly serve to mitigate risks. Therefore, the mere failure to fulfil these requirements such as log-keeping does not necessarily imply a violation of the duty of care under criminal law. Based on these observations, it can be argued that, in areas where the AI Act applies as an EU Regulation, the relevant obligations and requirements to mitigate risks of AI (-driven) systems may serve as a potential source of the duty of care.

It must be acknowledged that, for example, the requirements outlined in Article 8 and subsequent provisions concerning high-risk AI systems are specific to the AI Regulation. Compliance with these obligations and requirements alone does not eliminate the need to adhere to national legal prerequisites. For instance, in determining criminal liability in Germany, not only national regulations but also unwritten norms of conduct and the aforementioned sources must be taken into account. Nevertheless, the AI Regulation may exert an indirect influence on domestic legislation, requiring national criminal justice systems to adapt and incorporate clear and comprehensive provisions. Failure to implement such measures as prescribed could result in liability for negligence[1575].

As elaborated in detail above, compliance with such standards serves merely as an indicator for fulfilling the duty of care. Therefore, while adherence to these obligations and requirements will likely mean that the persons behind the machine have fulfilled their duty of care, this is not definitive. The general principle of refraining from causing harm remains applicable in all cases. Even the official approval of a product by the authority responsible for setting the legal framework to ensure safety, efficacy, and quality does not automatically release the manufacturer or seller from their duties[1576]. Thus, the AI Regulation's risk-acceptability threshold for particularly high-risk AI systems does not allow sole reliance on technical standards. Specifically, in situations where a reasonable provider could

[1575] ROMANO Leonardo, "Criminal negligence and acceptable risk in the EU's AI Act: casting light, leaving shadows", 24.09.2024; Lex ET Scientia International Journal (LESIJ), V. 1, I. 26, 2019, p. 146.

[1576] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1187, Rn. 280.

foresee that the system might cause harm[1577], mere compliance with the Regulation's provisions does not ensure the application of permissible risk. Indeed, it is highly problematic when large companies reduce their compliance efforts to a box-ticking exercise, merely meeting the standards on paper without substantive implementation[1578]. In every concrete case, whether the duty of care has been fulfilled must be carefully assessed in detail by the courts, and only when all relevant conditions are satisfied should the permissible risk doctrine be applied.

D. Criminal Liability Involving Multiple Actors and The Problem of Many Hands

1. The Concept of "the Problem of Many Hands"

The "problem of many hands," first introduced in 1980, refers to the challenge of attributing moral responsibility within complex organisational structures where numerous individuals contribute in varying capacities to decisions and policies. The involvement of multiple actors in such processes makes it difficult to determine who should bear moral responsibility for the outcomes[1579]. In situations where multiple individuals contribute to an outcome, the difficulty of identifying the morally responsible person has led some scholars to propose collective responsibility[1580]. However, such an approach is not feasible in the context of criminal liability.

In contemporary English-speaking legal literature, this concept is frequently employed in the assessment of legal and criminal responsibility. While it often arises in the context of product liability, its application is not limited to such matters; it is also relevant in determining responsibility within military settings[1581]. An example of the problem of many hands was in the 1980s, where the *Therac-25* radiation machine malfunctioned, overdosing six patients and causing three deaths. It occurred due to a combination of different factors: software errors, inadequate testing, poor

[1577] See: Chapter 4, Section C(3)(c): "Under Which Perspective Should the Standard of Care Established?".

[1578] ROMANO Leonardo, "Criminal negligence and acceptable risk in the EU's AI Act: casting light, leaving shadows", 24.09.2024.

[1579] THOMPSON, The Problem of Many Hands, 1980, pp. 905-916.

[1580] See: VAN DE POEL, The Problem of Many Hands, 2015, p. 55 ff.

[1581] NISSENBAUM, Accountability in a Computerized Society, 1996, p. 29.

design, and insufficient investigation. In a retrospective investigation it was not possible to blame a single person as multiple factors and actions contributed to the incidents[1582].

In this regard, the problem of many hands can be considered to have two dimensions in terms of causality and negligence. Accordingly, the aforementioned explanations are equally applicable in this context[1583]. AI-driven autonomous systems are developed through the involvement of numerous actors, both in terms of software and hardware. Consequently, attributing liability to a specific individual or group -such as those responsible for preparing the training dataset, designing parts of the machine learning algorithm, or contributing to the overall design- proves to be exceptionally challenging. This section will concentrate on the providing potential solutions for this issue, particularly within the context of the principle of reliance. However, the discussion will not be limited to this aspect alone; it will also seek to propose solutions to challenges that may arise from human-machine collaboration.

2. The Principle of Reliance

a. The Concept

The term *principle of reliance*[1584] is adopted in this study to refer to the concept of *Vertrauensgrundsatz* in German legal literature, because “principle of trust”[1585] does not sufficiently convey the essence of this principle. On the other hand, “reliance” more accurately reflects the legal context where parties act based on the reasonable expectations created by others, whereas “trust” is a broader concept that lacks this specific legal nuance.

[1582] NOORMAN Merel, “Computing and Moral Responsibility”, The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Eds.: Edward N. Zalta/Uri Nodelman, https://plato.stanford.edu/archives/spr2023/entries/computing-responsibility. (accessed on 01.08.2025).

[1583] See: Chapter 4, Section A: “Causality” and Chapter 4, Section C: “Negligent Liability”.

[1584] For an example of the use of the term *principle of reliance* in English literature, see: XU/HUANG, Traffic Crash Liability, 2016, p. 322.

[1585] For an example of the use of the term *principle of trust* in English literature, see: DUBBER/HÖRNLE, Criminal Law, 2014, p. 580.

According to a widely accepted view, the principle of reliance is a form of permissible risk[1586]. The principle of reliance in criminal law indicates that an individual who acts in accordance with legal rules may assume that others will also adhere to the law and act as law-abiding individuals. This principle allows individuals to base their actions on this reliance, without the need to constantly assess whether others are acting diligently or to adjust their behaviour to account for potential breaches of diligence. Thus, as a general rule, each person is responsible for their own conduct. However, the principle does not apply when there are clear and recognisable circumstances that undermine this reliance, such as situations requiring caution due to specific behavioural conditions that indicate that others may not act as expected[1587].

The principle of reliance initially emerged from the necessity of regulating traffic after rapid industrialisation and developed to address the practical demands of road safety. In this context, individuals needed to rely on the predictable and responsible behaviour of others to ensure orderly and secure traffic flow. However, over time, the principle evolved beyond its origins in traffic law and extended into broader legal contexts[1588]. This development can be attributed to the growing importance of the division of labour and specialisation, both of which require individuals to rely on the competence and diligence of others[1589].

In the assessment of negligence, the principle of reliance establishes that causal outcomes arising from situations in which the perpetrator can

[1586] WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 824, Rn. 92; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 319 Rn. 823; AKBULUT, Ceza Hukuku, 2022, p. 410.

[1587] WELZEL, Das deutsche Strafrecht, 1969, p. 133; VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1165 f, Rn. 229; RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 534 Rn. 22 f.; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 186 f. Rn. 61 ff.; KATOĞLU, Ekip Halinde, 2007, p. 31 f.; EIDAM, Zum Ausschluss, 2011, p. 913;

[1588] AKBULUT, Ceza Hukuku, 2022, p. 411.

[1589] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1166, Rn. 232; KATOĞLU, Ekip Halinde, 2007, p. 32.
The principle of reliance was gradually adopted by legal systems; for instance, in Italy, the Court of Cassation initially refused to recognise the preventive effect of the principle of reliance in negligent liability in traffic cases. See: DELOGU, Modern, 1987, p. 124.
For an analysis of certain decisions of the Court of Cassation, see: KATOĞLU, Ekip Halinde, 2007, p. 34.

rightfully rely on that a certain event will not occur (particularly in relation to the conduct of third parties), cannot be objectively imputed to the perpetrator, provided that there is no breach of the duty of care[1590]. In this regard, the principle of reliance also serves to impose a limit on the objective duty of care[1591].

By its nature, complicity in negligent offences is not possible[1592]. Thus, the concept is closely connected to the principle of individual criminal responsibility, whereby individuals are liable solely for their own behaviour and cannot be punished for the conduct of others. Accordingly, every individual need only comply with the norms of conduct that concern their own behaviour[1593]. In this regard, according to this principle, the limits of careful or permissible risky behaviour should, in principle, be determined without taking into account the potential misconduct of others. It is also to be assumed that others will act with due care and within the bounds of permissible risk[1594].

Although common experience suggests that others involved in a harmful outcome often act negligently, a person is not always required to adjust their behaviour to prevent the harm caused by the negligent behaviour of others and can reasonably rely on the expectation that others will fulfil their own duties of care[1595]. In this way, for example, a driver approaching an intersection on a public road is not expected to completely stop and meticulously check the road to eliminate all possible risks. Instead, the driver may proceed through the intersection (where they have the right of way) by reasonably slowing down. If, as a result, another vehicle collides with them, the liability lies with the driver who caused the collision. Indeed, without the principle of reliance, it would be nearly impossible to maintain normal and smooth traffic flow due to the excessive liability risks that could arise[1596].

According to the German Federal Court of Justice (BGH), the principle of reliance also applies in other areas where multiple individuals work

[1590] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 186 Rn. 61.

[1591] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 263 Rn. 26.

[1592] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1104.

[1593] KATOĞLU, Ekip Halinde, 2007, p. 31 f.

[1594] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1163 f., Rn. 224.

[1595] PUPPE, § 5 Der Vertrauensgrundsatz in Strafrecht AT, 2023, p. 89 Rn. 21.

[1596] HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 559; DOĞAN, Sürücüsüz Araçlar, 2019, p. 3232.

together in a division of labour. For instance, an anaesthetist may rely on the surgeon to properly coordinate their activities with those of the anaesthetist[1597]. However, whether the principle of reliance can be invoked must be determined separately in each individual case, as the boundaries of the division of labour are often not clearly defined[1598].

Another example can be given where a customer dies as a result of a meal served by the waiter who did not know it was poisoned. The waiter cannot be held liable even if they hated the customer and wished for their death one day; unless it could be foreseen that the food was poisoned, such as the cook being capable of such behaviour[1599]. As this example demonstrates, the principle of reliance has its limits.

The principle of reliance in criminal law is no longer applicable when it becomes evident that (through concrete indications) it is unreasonable to expect proper or lawful behaviour from others, or when the actor is aware -or ought to be aware- of circumstances that make noncompliance foreseeable and preventable[1600]. In such cases, if a danger has already arisen due to another's negligent conduct[1601], or if a person occupies a position of hierarchical or legal authority that imposes a duty of supervision and intervention, any reliance on the adherence of others to rules is displaced by the necessity to anticipate and avert harm[1602]. Similarly, when there are evident indications that another party is behaving improperly, is evidently incapable of adhering to the rules (for instance, due to intoxication or inexperience), or is likely to violate safety norms based on recognisable tendencies of misconduct, the actor cannot invoke the principle of reliance merely by fulfilling their own responsibilities. Therefore, once it becomes evident that reliance on another's compliance is no longer reasonable, the principle of reliance is replaced by the obligations of foresight, diligence, and the

[1597] Federal Court of Justice (BGH), judgment of 02.10.1979, Case No. 1 StR 440/79, reported in NJW 1980, p. 650.

[1598] KATOĞLU, Ekip Halinde, 2007, p. 35.

[1599] DUTTGE, Erlaubtes Risiko, 2010, p. 146.

[1600] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 263 Rn. 26; HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 559; STRATENWERTH, Zur Individualisierung, 1985, p. 301; HEGER, StGB § 15 in StGB Kommentar, 2023, Rn. 39a.

[1601] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 320 f. Rn. 64.

[1602] KATOĞLU, Ekip Halinde, 2007, p. 32, 35-36; AKBULUT, Ceza Hukuku, 2022, p. 412.

proactive avoidance of foreseeable harm, where applicable[1603]. However, if the perpetrator cannot recognise this fact, it can be taken into account[1604]. Nevertheless, one cannot rely on others to compensate for dangers they have created through their own negligent behaviour or violation of safety rules, as the principle of reliance does not protect those who neglect due care or established safeguards[1605].

b. The Problem of Many Hands and AI-Driven Autonomous Systems

Addressing the "problem of many hands" becomes particularly complex when multiple actors contribute to a harmful outcome in diverse ways and to varying degrees. In such cases, where a product is involved, the established mechanisms of criminal product liability are generally applicable. Nevertheless, adding to this complexity, the opacity of AI-driven autonomous systems, as discussed in detail above[1606], particularly the issue of the *black-box* nature of such systems, aggravates the difficulty of resolving liability. In such cases, the inability to determine whether the harm originates from training data, flawed programming, a system bug, or a combination of these factors[1607] makes it nearly impossible to ascertain which actor contributed to the outcome and in which manner[1608]. Consequently, attributing liability to a specific individual becomes practically unattainable[1609].

This problem arises not only in instances involving the failure of a single AI-driven autonomous system; but also in scenarios where multiple

[1603] VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1165, Rn. 227; KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 187 Rn. 63 ff.; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 303 f. Rn. 40; GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 563 Rn. 62; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 225 Rn. 31; TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 238; KATOĞLU, Ekip Halinde, 2007, p. 34; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1121.

[1604] PUPPE, § 5 Der Vertrauensgrundsatz in Strafrecht AT, 2023, p. 84 Rn. 8.

[1605] STRATENWERTH/KUHLEN, § 15 Das fahrlässige in Strafrecht AT, 2011., p. 321 f. Rn. 67.

[1606] See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".

[1607] COOPER, et al., Accountability, 2022, p. 864 ff.

[1608] See: Chapter 4, Section C(4)(b)(1): "The Anatomy of Failures in AI-Driven Systems".

[1609] COOPER, et al., Accountability, 2022, p. 866 ff.

systems interact with each other and with humans in their environment, potentially causing harm. In such situations, the difficulty of assigning liability is further complicated. According to one perspective, when numerous unpredictable AI-driven systems act as collaborators in causing harm, the traditional principle of reliance may prove insufficient and may require reconsideration[1610]. Moreover, an inadequately designed liability regime could result in both liability gaps and overlapping liabilities[1611].

(1) Liability Challenges in the Production Chain of AI-Driven Autonomous Systems

Sole ownership businesses, once common, have become increasingly rare as modern production and distribution companies predominantly adopt corporate structures to accommodate the complexity and scale of contemporary business operations[1612]. For instance, even software development has long been a collaborative effort, bringing together individuals from diverse fields; such as designers, engineers, programmers, graphic designers, managers, and others to create a final product. However, despite the inherently collective nature of such processes, the concept of liability, particularly in criminal law, centre the individuals[1613]. As highlighted in discussions on product liability[1614], determining which actor's behaviour led to a harmful outcome becomes particularly challenging when multiple actors are involved in the production process, such as in the creation of software and hardware[1615].

Due to the complexity of modern production processes, it is rarely feasible to identify a single individual who is solely responsible for the harmful outcome, especially when employees operate within complex collaborative systems[1616]. This difficulty is further impaired in cases involving AI-driven bots and robots, where the hardware components and software elements

[1610] KAIAFA-GBANDI, Artificial intelligence, 2020, p. 323.
[1611] NOVELLI/TADDEO/FLORIDI, "Accountability in AI, 2023, p. 5.
[1612] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1938.
[1613] NISSENBAUM, Accountability in a Computerized Society, 1996, p. 29.
[1614] See: Chapter 4, Section C(1)(d): "Product Liability".
[1615] OSMANI, The Complexity of Criminal Liability, 2020, p. 65.
[1616] HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 448.

may be produced by different manufacturers. Such fragmentation complicates the identification of the specific cause of a failure[1617].

When an AI-driven autonomous system is involved in or causes a criminal offence due to a failure, the failure can arise from a variety of causes. It may result from a defect in the software or hardware, an error attributable to the human operator, or issues stemming from the system's operation within real-world parameters, particularly in the context of unexpected events. Moreover, it is likely that such failures arise from a combination of these factors. Indeed, even under normal circumstances, identifying problems in software and hardware is inherently challenging[1618]. Furthermore, on the one hand, the complexity of AI systems is desirable as it enhances the system's performance based on the chosen model. On the other hand, this very complexity and opacity makes it significantly more difficult to establish causal relationships during *ex post* assessments[1619].

Detecting software-related issues is particularly challenging. This is partly due to the fact that different individuals are typically responsible for various components of the software, and also because software is rarely developed entirely from scratch. Instead, it is often built in combining with or atop other software, which requires compatibility and integration. Algorithmic systems that process data frequently rely on toolkits developed externally, which may already have inherent issues. Furthermore, machine learning toolkits often incorporate extensive, pre-trained models, adding another layer of complexity to pinpointing the exact cause of a problem. Issues may arise from the training data itself, even in its filtered form, or from a misalignment between hardware and software. In the context of AI systems, these challenges are magnified, as some components may be outsourced or obtained from third parties[1620].

Each issue that may arise from these components can be linked to the specific processes within the collaborative endeavour of AI development. The involvement of diverse teams of programmers and specialists in developing AI systems complicates the identification of, for instance, the specific programmer responsible for the line of code that triggered the system's con-

[1617] BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 5.

[1618] GOGARTY/HAGGER, The Laws of Man over Vehicles Unmanned, 2008, p. 73.

[1619] BECK, Google Cars, 2017, p. 243.
See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".

[1620] NISSENBAUM, Accountability in a Computerized Society, 1996, p. 29 f.; COOPER, et al., Accountability, 2022, p. 867 f.

duct[1621]. Moreover, this often does not stem from a single cause. Challenges may also emerge during the development phase as a result of unintended consequences stemming from decisions made by key actors. Furthermore, hierarchical organisational structures can inadvertently contribute to these challenges, particularly when individuals who are not directly involved in specific tasks influence critical decisions[1622]. In fact, a self-driving vehicle accident might result from a combination of general factors, such as misconduct by data labellers, careless oversight by a programmer or quality control staff, a mechanical defect in the vehicle's sensors, and indirectly the managing board's prioritisation of quick profit over thorough evaluation[1623].

To illustrate, it is almost impossible for a company manufacturing self-driving vehicles to design and produce all components -such as sensors, cameras, batteries, LIDAR, radar, complete software systems, and other mechanical parts- entirely within its own organisation, as each requires specialised expertise. However, when a self-driving vehicle is involved in an accident, the issue could stem from any of these components or, alternatively, from the software, such as a failure in the image recognition system; or from the interaction between these components as well as their failure to function harmoniously. In such cases, identifying the specific cause becomes exceedingly difficult. In cases of hardware failure, for instance, if the company provides its chips from another supplier, it is generally entitled to rely on the assumption that the chips are free from defects, provided that they have undergone reasonable testing. The company cannot be expected to check every chip as if they were the manufacturer, especially considering that they may lack the technological capacity to do so. Nonetheless, releasing the final product into the market without conducting any inspection would constitute a breach of their duty of care. Here, the principle of reliance applies; however, the company retains a duty of control, which varies depending on the degree of risk involved and the legal interests at stake.

A clear example of this issue is the 2016 fatal Tesla accident discussed earlier, where one of the contributing factors was the integration of a front-facing camera sourced from another company into Tesla vehicles. The resulting fatality raises a challenging question: could Tesla's officials reason-

[1621] VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 665.

[1622] NISSENBAUM, Accountability in a Computerized Society, 1996, p. 29.

[1623] GIANNINI/KWIK, Negligence Failures, 2023, p. 59.

ably rely on the other company, given the compatibility issues between the camera and the vehicle? In this context, both companies have essential responsibilities, but eventually it was Tesla's responsibility to conduct the necessary testing. Another example of an accident resulting from the combination of multiple factors is the 2018 Uber crash discussed above. In this case, the collision occurred due to a combination of the test driver's inattention, errors in the vehicle's software, and the pedestrian's own lack of caution, ultimately resulting in a fatality[1624].

In cases where a harmful outcome arises from a company's product, it is logical to begin the analysis of a potential breach of the duty of care by examining the company's organisational structure. This is because every company's hierarchical setup differs, with varying allocations of oversight responsibilities and relational networks among its management. In such instances, the internal distribution of responsibilities must be identified and assessed in the context of the specific case[1625]. In line with the principle of reliance, the necessity of trust in cooperative endeavours, particularly those reliant on a division of labour, combined with the complexity inherent in technical contexts, limits the extent to which individuals can be held liable for collectively caused damages[1626].

In organisations such as companies, the division of labour can be distributed both horizontally and vertically. A horizontal division of labour refers to a collaborative process where multiple individuals of equal status perform different tasks simultaneously within a shared project or system. The principle of reliance does not apply when there are clear signs that one of the collaborators is acting in a way that is evidently faulty or poses an obvious risk to the outcome[1627]. On the other hand, a vertical division of labour refers to the hierarchical distribution of tasks within a professional field, where responsibilities are delegated from a superior (such as a chief physician) to subordinates (doctors and non-medical staff). This structure is based on reliance, with the chief responsible for overseeing tasks and

[1624] See: Chapter 2, Section C: "Prominent Cases Highlighting AI-Related Liability".

[1625] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 175.

[1626] IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 429.

[1627] EIDAM, Zum Ausschluss, 2011, p. 914.
For instance, significant emphasis is placed on the duty of supervision and care in the field of occupational health and safety in Turkish jurisprudence. The Court of Cassation, in a case, has held employers liable for breaching their duty of supervision and oversight as they failed to employ qualified workers in hazardous areas of the workplace. For the assessment, see: KATOĞLU, Ekip Halinde, 2007, p. 35 f.

subordinates following instructions, but both parties may bear liability depending on their adherence to delegated duties and instructions. Subordinates are personally liable when performing tasks independently, and the superior can rely on their proper execution if they have selected, instructed, and organised their staff and processes appropriately[1628].

In a division of labour, every diligent member of an organisation may reasonably rely on others to perform their tasks with due care, unless there are clear indications that the principle of reliance does not apply, such as evidence that the other party is failing to fulfil their duty of care[1629]. However, in many cases, the outcome arises from the involvement of multiple individuals, making it possible that ultimately no one can be held accountable for the result[1630]. Alternatively, when one party's act in violation of due care is combined with a similar act by another, the outcome can be objectively imputed to all involved. In such cases, responsibility may not rest with a single individual; rather, each party can be separately held liable in accordance with their negligent behaviour. The key condition for such attribution is that all liable individuals must have breached their duty of care[1631]. The primary issue arises in situations where none of the individual actions can be characterised as a breach of the duty of care, yet their cumulative effect results in a harmful outcome.

In hierarchical structures, the principle of reliance may, in certain circumstances, relieve a superior of liability by allowing them to rely on employees to act prudently. However, this presumes that the superior has fulfilled their duties of care, which extend beyond selecting a professionally and personally suitable individual among applicants to include proper guidance and supervision. When these obligations are met, the superior may generally rely on the fact that subordinates will perform their tasks appropriately[1632]. Nevertheless, such vertical divisions of labour do not create entirely divided responsibilities or liabilities; instead, they result in overlapping and multiplied individual responsibilities[1633]. Furthermore, the principle of reliance does not apply in cases where the duty of care specifi-

[1628] EIDAM, Zum Ausschluss, 2011, p. 915.

[1629] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 9.

[1630] *Ibid.*

[1631] KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, p. 224; DEMIREL, Otonom, 2024, p. 1262.

[1632] GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 564 Rn. 65; ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 180.

[1633] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 176.

cally entails preventing the misconduct of third parties, such as within the scope of control and supervisory duties[1634]. This individual is responsible for both their own tasks and overseeing the work of others as part of the division of labour. However, the duty of supervision and control cannot be unlimited, as its purpose is not to designate a single person at the top as liable in every situation[1635].

In certain cases, business areas within a management board may be divided based on areas of expertise or specific roles, such as a deputy managing director responsible for a particular field. If the outcome arises from an issue within that specific area, as a rule the relevant managing director should be held liable[1636]. However, the concept of general responsibility can be seen in the German Federal Court of Justice (BGH)'s *Lederspray* decision[1637] which demonstrates that the division of business areas among directors does not absolve any individual director from responsibility for the overall management of the company. Under this principle, every board-member who is responsible for the decisions of the company in general is required to ensure legal compliance, even when tasks are delegated or specialised. While reliance on the expertise of colleagues is permitted, board members have a duty to intervene when risks are apparent and cannot evade liability through the division of business. Ultimately, it does not result in a collective criminal liability; it is assessed individually, based on what each director knew, ought to have known, and the reasonable steps they took to prevent the harm[1638].

The division of labour within a company does not diminish individual responsibility; rather, it multiplies it, as overlapping duties and the complexity of organisational structures can result in multiple employees being held criminally liable for the same incident[1639]. In cases involving product defects, current criminal law tools can generally identify the responsible parties[1640]. However, when it comes to AI-driven autonomous systems, particularly that continue to learn after being deployed, identifying responsible

[1634] KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 225 Rn. 32.

[1635] DEMIREL, Taksir, 2024, p. 300 f.

[1636] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1940.

[1637] See: Chapter 3, Section C(1)(d)(6)(c): "Key Judicial Decisions Shaping Criminal Product Liability".

[1638] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung Das Lederspray-Urteil des BGH, 1990, p. 2966, 2969; KUHLEN, Grundfragen, 1994, p. 1145 ff.

[1639] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1942.

[1640] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 177.

actors becomes nearly impossible. Those involved in the production of such systems must exercise the utmost care. According to one perspective, in such cases, the responsibility for preventing harmful outcomes does not rest solely on the manufacturers; it is shared with buyers, trainers, and all parties involved in deploying the systems[1641].

If a product is prematurely released on the market, responsibility initially falls within the internal corporate domain of the individual overseeing the relevant department, such as development or production management, depending on where the failure or oversight occurred. This aligns with the principle that criminal liability in such cases depends on identifying the individual within the organisation who had the specific legal duty to prevent the harmful outcome[1642]. In particular, during crises or exceptional situations requiring a product recall, ultimate responsibility reverts to superior management[1643]. Criminal liability for breaches of company-related duties of care is not confined to the individual directly responsible; it may extend to superiors, colleagues, or employees who share responsibility due to their organisational, supervisory, or reporting obligations[1644]. According to one perspective, in the event of an incorrect majority decision within a collegial body, the potentially responsible individual, in fulfilling their duty of care, must advocate for the correct decision, report the issue to higher management, and, if the risk is significant, even make the matter public[1645]. Under this view, an employee who identifies a potential problem and reports it to their hierarchical superior should not be held liable if the offence subsequently occurs[1646].

(2) Other Instances of the "Problem of Many Hands" in Relation to AI-Driven Autonomous Systems

The potential involvement of multiple actors in situations where AI-driven autonomous systems are implicated in a criminal offence is not limited to the production chain. The problem of many hands in relation to such autonomous bots or robots may arise from a variety of scenarios involving

[1641] WOLF/MILLER/GRODZINSKY, Why We Should Have Seen That Coming, 2017 p. 2 f.

[1642] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1938.

[1643] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 176.

[1644] SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1939.

[1645] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 181.

[1646] MÜSLÜM, Artificial Intelligence, 2023, p. 142.

their interaction with the environment. For instance, questions such as whether it is legally reasonable for self-driving vehicles to rely on the assumption that a pedestrian will not suddenly step onto the road will be addressed below. Nonetheless, it should be stated that, in situations involving the use of AI-driven autonomous systems where multiple individuals are involved, the principle of reliance is applied to the extent that it aligns with its inherent nature and purpose.

In cases where an AI system is developed within an organisation such as a company, despite various challenges, it is at least possible to retrospectively identify errors made by a specific developer in a portion of the code through tools such as '*git blame*'[1647]. However, the situation is far more complex for AI systems developed using *open-source software*[1648]. In my view, the applicability of the principle of reliance in this context is limited; developers have a greater obligation to review and verify the contributions of their predecessors. This is because, in the absence of a structured division of labour among contributors, a higher standard is required for reliance to be deemed reasonable. In open-source software, the source code is made publicly available under the terms of an open-source license, allowing anyone to use, modify, or distribute it. Numerous individual developers contribute to the code in diverse ways, often making their work available for further use by others. Unlike in a corporate setting, where developers work within a structured framework, these individuals operate independently. Consequently, it can be argued that the individual who will use the final system bears the responsibility to thoroughly review the entire system. It can also be stated that the duty of care intensifies in accordance with the nature of the work performed.

A similar issue may arise when a company developing for instance, a large language model (LLM) provides APIs[1649] to other developers, enabling them to customise the model for specific personal or professional uses. In such scenarios, determining whether a harmful outcome (such as

[1647] *Git blame* identifies the author and details linked to each line in a file, thus enables the tracing of changes and their origins.

[1648] For instance, pursuant to Article 2(2), the new Product Liability Directive does not apply to free and open-source software that is developed or supplied outside the scope of a commercial activity.

[1649] API (Application Programming Interface) is "a set of rules or protocols that enables software applications to communicate with each other to exchange data, features and functionality". GOODWIN Michael, "What is an API (application programming interface)?", 09.04.2024, https://www.ibm.com/think/topics/api. (accessed on 01.08.2025).

the model insulting users during its operation) originates from the original product or the customised version can be challenging. If both the original developers and those customising the system have breached their respective duties of care, none are exonerated, even if the harm could have been avoided by the diligent conduct of just one actor. This is rooted in the principle of victim protection, which ensures that no party can evade liability by claiming that the other party's individual care alone would have prevented the harm[1650].

Another problem of many hands related to AI-driven autonomous systems arises in scenarios where the harmful outcome results from the added faulty behaviour or assumption of risk by third parties. Ordinarily, the required level of care is limited by the principle of reliance, which presumes that others will act responsibly and with due care[1651]. However, if the risks associated with an AI-driven autonomous system are well-known, the system does not guarantee absolute safety, and the manufacturer has provided clear warnings about clear and potential dangers; a person who chooses to implement the system despite these warnings is considered to have assumed the risk[1652]. Assumption of risk differs from consent as the injured party retains control over the damaging causal process, knowingly engaging with the situation despite awareness of the potential hazards[1653]. On the other hand, if the offence occurs due to the victim's creation of the risk, and they act on their own responsibility, the perpetrator cannot be objectively imputed with liability in such a case[1654]. However, in a case where both the perpetrator and the victim has violated due care, and the victim's careless behaviour is substantially less relevant than the perpetrator's in causing the harmful outcome, the perpetrator's liability for negligence persists[1655].

[1650] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 188 f. Rn. 74.

[1651] HILGENDORF, Moderne Technik, 2015, p. 101.

[1652] SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 501.

[1653] KINDHÄUSER, Zum sog. 'unerlaubten' Risiko, 2010, p. 415.

[1654] KINDHÄUSER/ZIMMERMANN, § 11 Objektive Zurechnung beim Erfolgsdelikt: Strafrecht AT, 2024, p. 107 Rn. 24.

[1655] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1135.

c. Introducing AI-Driven Autonomous Systems into the Principle of Reliance

As humans and machines increasingly collaborate in daily life, autonomous systems have begun to take on certain tasks that were conventionally performed by humans, demonstrating capabilities that closely mimic human-like functionality. This shift has sparked debates about whether the principle of reliance, which allows individuals in a division of labour to rely on the assumption that others will comply with the law and act as responsible participants, can be extended to include AI-driven autonomous systems. The question here is whether humans can rely on autonomous and fully automated systems to function correctly and whether these machines (autonomous systems) should take human error into account[1656]. Naturally, this leads to a further question: should humans instead act with constant readiness for potential errors by such systems? Furthermore, another question that needs clarification is whether the reliance is placed on the person behind the machine or on the machine itself. Additionally, must these systems be legally classified as an actor or agent to be included under the principle of reliance?

Indeed, the principle of reliance is already applied to conventional vehicles and, with certain limitations, also governs interactions between the driver and the system[1657]. In tasks involving collaboration between humans and machines, the concept of the human-machine interface is frequently discussed. Accordingly, clear communication and effective transfer of responsibility between the human and the machine are essential to ensure that both parties are fully "aware" of their roles during control transitions, thereby preventing harmful outcomes[1658]. However, risks may increase in such scenarios. For instance, humans may become less cautious in certain tasks, presuming that autonomous systems will compensate for their lack of attention or carelessness. Therefore, liability rules must be designed comprehensively to ensure that no gaps are left in addressing such situations[1659].

[1656] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, pp. 11-12.

[1657] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 214.

[1658] *Ibid*, p. 70 f.

[1659] DI/CHEN/TALLEY, Liability Design, 2020, p. 3.

(1) Should Humans Rely on Machines?

As human-made systems increasingly take over certain tasks and as the testing processes in their development become more rigorous to ensure their safety and reliability, greater trust is placed in these systems. This trust largely stems from the expectation that the system will perform the assigned task as anticipated, in the expected manner, and within the expected timeframe. In this context, it can be argued that trusting these systems on the presumption that they will function reliably, is reasonable; particularly when they meet or exceed the standards expected of humans, whose error rates are typically higher due to external factors such as physical and emotional conditions[1660]. Indeed, it is generally accepted that in autonomous systems, as users place greater trust in the technology, responsibility tends to shift more significantly toward the manufacturers[1661].

Thus, it is argued that the principle of reliance can be extended to human-machine interactions, on the premise that AI-driven autonomous systems such as self-driving vehicles will adhere to established regulations and incorporate appropriate technical safeguards. However, this principle is applicable only in the absence of clear indications of malfunction. If warnings from manufacturers, media reports, or observable anomalies in the system's conduct suggest potential issues, the principle of reliance ceases to apply[1662].

It is reasonable to rely on an automated or AI-driven autonomous system to function correctly in the future if it has consistently operated properly in the past. This reliance is particularly acceptable given the growing prevalence of complex technological devices, which are replacing simpler tools. The reliable and consistent functioning of these advanced systems fosters confidence in their proper operation. Indeed, it is impractical in daily life to inspect every component of such systems in meticulous detail. For instance, while an individual may check their car tyres regularly before travelling; inspecting the engine, brakes, and other components daily would be incompatible with the ordinary course of life. At some point, reliance on

[1660] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 169.

[1661] SEUFERT, Wer fährt, 2022, p. 321; BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 8

[1662] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 11 f.; HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 453; HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 154; HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 559.

the assumption that these parts will function properly becomes a practical necessity.

However, this reliance must have its limits. Blindly adhering to the outputs of a system that has produced accurate results in the past, without questioning its future outputs, may lead to *automation bias*[1663] and result in reducing their level of active engagement and eventually a failure to exercise due care[1664]. In this regard, the question should be answered: can the operator be accused of negligence in an incident due to a malfunction if the autonomous system has always worked faultlessly in the past[1665]? The level of reliance placed in autonomy must not be overestimated, ensuring that the required standard of care and diligence is not diminished[1666]. For instance, a driver must not place blind trust in a navigation device; instead, they should exercise their own judgment and conduct necessary checks to ensure careful and responsible use[1667].

Nevertheless, the answer may not be straightforward. For instance, in a scenario where a driver, acting on a navigation device's instruction to "turn right" in foggy conditions, follows the directive and ends up driving into a river, causing both themselves and a passenger to drown, a question arises: could the manufacturers of the navigation system be held liable for such outcome by negligence? Such questions can be multiplied. For instance, if the passenger, rather than the driver who trusted the system and assumed the risk, had drowned, who should be held liable? Alternatively, what if someone in the front passenger seat had been giving directions and provided incorrect guidance? Or, what if the driver had been navigating using a printed map that contained an error, leading to the vehicle's being driven into the river[1668]? In such cases, individuals must verify the navigation system's instructions before acting on them; otherwise, they cannot

[1663] Automation bias is a decision-making phenomenon where humans have a tendency to disregard or not search for contradictory information in light of a computer-generated solution that is accepted as correct. See: CUMMINGS, Automation Bias, 2004, p. 2.

[1664] GIANNINI/KWIK, Negligence Failures, 2023, p. 73 f.; SMILEY Lauren, "'I'm the Operator': The Aftermath of a Self-Driving Tragedy", 08.03.2022, https://www.wired.com/story/uber-self-driving-car-fatal-crash. (accessed on 01.08.2025).

[1665] HILGENDORF, Grundfragen, 2013, p. 27.

[1666] PEKMEZ KELEP, Otonom Araç, 2018, p. 174 f.

[1667] SCHUSTER, Providerhaftung, 2017, p. 56.

[1668] JOERDEN, Strafrechtliche Perspektiven, 2013, pp. 195-196.

evade liability. That said, it is argued that the liability of navigation system's programmer can be discussed[1669].

This question becomes more complex when a human driver is replaced by an autonomous system. For instance, if a vehicle were under the control of a self-driving system which followed the instructions of a navigation system outsourced from another company, leading to the vehicle plunging into a river, how would liability be determined? In my view, a self-driving vehicle must rely on its own sensors to perceive its surroundings and act accordingly, rather than placing unconditional trust in data from a single source, such as a navigation system. This conclusion can be reached based on various general principles. Nonetheless, considering the current functionality of navigation systems and the level of reliance placed in them, it can be observed that they have evolved beyond merely serving as auxiliary tools for obtaining guidance.

As observed previously, assessing humans' trust in machines under the principle of reliance appears challenging in the present context. Beyond its theoretical challenges, particularly in today's transitional phase, individuals are expressly burdened with a duty of care that includes the obligation to verify the proper functioning of these systems[1670]. Moreover, according to one view, applying the principle of reliance in human-machine and machine-machine interactions is currently not feasible, as it is not yet fully possible to anticipate the conduct of such systems. They are considered unpredictable for humans and are not governed by reason; which makes them a source of danger rather than a reliable agent[1671]. Therefore, they conflict with the norms and expectations governing human interactions[1672]. Furthermore, with autonomous vehicles and interconnected driving systems, it becomes nearly impossible to ascertain who (a human or a self-driving system) is operating another vehicle and on what basis they are making their decisions[1673].

Another criticism can be raised regarding which machines should be included under the principle of reliance? For instance, should complex systems like self-driving vehicles be included, while systems consisting solely of software, such as LLM chatbots, are excluded? What about simpler

[1669] *Ibid*, p. 206.

[1670] See: Chapter 4, Section C(4)(d): "Control Dilemma".

[1671] FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 886.

[1672] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 445 f. Rn. 27; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 169 f.

[1673] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 450 Rn. 41.

systems like internet cookies? The question of where to draw the line inevitably arises, which needs to be addressed. It can further be argued that automated systems are more predictable and, consequently, more reliable than autonomous systems. In this context, could a simpler system, such as a barrier that opens upon scanning a card, be evaluated within this scope?

(2) Should Autonomous Systems Rely on Humans?

Another issue concerning the principle of reliance is whether the design of AI-driven autonomous systems must account for human error, or whether these systems can be developed on the assumption that others (such as road users, whether human or even other self-driving vehicles) will behave in compliance with the rules[1674]. The 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? Moreover, if the manufacturer was in a position to foresee and prevent a common and identifiable human error, yet the autonomous system failed in this regard, can the manufacturer be held liable for such failure?

To concretise this question within the context of road traffic, a self-driving vehicle lawfully operating on the road detects, through its camera and LIDAR systems, a person preparing to cross the street at a red light. However, traffic continues to flow, and the vehicle relies on the assumption that the individual will not step onto the road against the light. Should the vehicle, in such circumstances, continue driving without reducing its speed, trusting that the individual will not act unpredictably? If the individual unexpectedly steps onto the road, causing an accident, should the liability of the person behind the machine be subject to legal examination?

An illustrative example is the 2017 media coverage of an incident where a robot allegedly saved a child who was climbing onto a toppling shelf by stabilising it. Although the event did not actually occur as reported and was misunderstood, it nonetheless serves as a good example for the purposes of

[1674] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 806.

this analysis[1675]. In incidents of this nature, the purpose for which the robot is deployed and its standard conduct must be considered, particularly in intersection to instances of human misbehaviour. For example, robots may potentially be utilised in childcare in the future. If robots are produced with the specific promise of supervising children within a certain age group, they must account for scenarios such as children climbing on shelves. This is because, above all, children's behaviour is inherently unpredictable, and such contingencies should be addressed when these robots are deployed in accordance with the promises made regarding their functionality.

Various examples can be provided on this subject. Undoubtedly, this issue holds significant importance for developers who create AI (-driven) systems and make them available for use by others. For instance, should manufacturers who produce an AI (-driven) system and make it publicly accessible online take precautions against potential misuse by third parties for purposes such as financial manipulation? Alternatively, can it be categorically argued that these systems are neutral by their dual-use nature, absolving developers of liability for their misuse? A pertinent example in this context would be whether OpenAI could bear responsibility if a third party misuses ChatGPT's API access for unlawful purposes.

In my view, rather than providing a direct answer to this question, it would be more appropriate to approach the matter in a nuanced manner. This requires a thorough examination of the issue within the framework of existing debates in criminal law dogmatics, particularly by considering the prohibition of regression (*Regressverbot*).

It should first be stated that, if no risk-indicating circumstances were recognisable *ex ante*, the subsequent chain of events would not have been foreseeable. For example, in a case where the perpetrator injures the victim due to excessive speeding but the victim subsequently dies in a hospital fire, according to one perspective, the occurrence of the fire is not a realisation of the risk created by the speeding. This sequence of events represents an unforeseeable circumstance. In this scenario, the risk of death

[1675] "Astonishing moment a ROBOT 'saves a girl from being crushed': Manufacturers claim machine moved forward and raised its arm to stop shelves toppling onto child 'despite NOT being programmed to do that'", 06.07.2017, https://www.dailymail.co.uk/news/article-4670544/Russian-robot-saves-girl-crushed.html. (accessed on 01.08.2025).

from excessive speed did not materialise. Therefore, causation (or objective imputation, according to the view adopted) cannot be established[1676].

The principle of reliance applies, in principle, so that one can rely on others not committing intentional crimes, including the sale of potentially dangerous products, since modern social life would be impossible if one had to constantly anticipate misuse for criminal purposes[1677]. In this regard, an individual who sells, lends or leaves lying around dangerous objects (axes, knives, matches, etc.) with which third parties could commit intentional crimes may reasonably rely on the presumption that no such acts will occur. However, the principle of reliance no longer applies if there are (concrete) indications to undermine this reliance[1678] or when one's actions encourage the apparent criminal intent of a potential perpetrator[1679].

For instance, a police officer places their gun on the table upon returning home. Their spouse, who has been waiting for an opportunity to kill a neighbour, takes the gun and commits the murder. In this scenario, there is an undeniable causal nexus. In this regard, since complicity is not present in the incident, the question arises as to whether an intentional or negligent act that follows the police officer's negligent behaviour can be attributed to them[1680].

According to the *prohibition of regression* (*Regressverbot*), the intentional action of another person is regarded as an intervening cause[1681]. However, according to principle of reliance, the nature and extent of one's duty of care depend also on the objective likelihood of the danger being exploited by third parties. Objects which typically pose a danger to the legal rights of others, even when used properly, require particularly careful safeguarding. There may be explicit legal provisions regulating such dangerous objects.

[1676] KINDHÄUSER/HILGENDORF, §15 Vorsätzliches und fahrlässiges Handeln - Strafgesetzbuch, 2022, p. 184 f. Rn. 55; KINDHÄUSER/ZIMMERMANN, § 33 Fahrlässigkeit - Strafrecht AT, 2024, p. 301 Rn. 30.

[1677] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1193 Rn. 26.

[1678] RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 546 Rn. 58.

[1679] ROXIN/GRECO, § 24. Fahrlässigkeit in Strafrecht AT, 2020, p. 1193 f. Rn. 28.

[1680] For the example, see: HAKERI, Ceza Hukuku, 2022, p. 192.

[1681] It is stated that a prohibition of regression -which would preclude prior negligent behaviour by the perpetrator or a third party that enabled an intentional act from being considered as a basis for causality- is not recognised by the prevailing opinion, because it cannot be explained by the condition theory and the equivalence of all causes. However, an interruption of the chain of attribution is conceivable. See: WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 244.

For example, Section 14(2), Sentence 2 of the StVO stipulates that motor vehicles must be secured against unauthorised use. Such provisions aim to mitigate the risks associated with the misuse of inherently dangerous items by imposing specific duties of care on their owners or users. On this matter, the OLG Stuttgart made significant determinations in its judgement concerning an arsonist who set a building on fire by misusing a landlord's temporarily stored waste[1682]. Accordingly, when inherently dangerous or easily misused objects are not secured as specifically legally required as such, and third parties misuse them to commit a negligent or intentional crime due to this lack of security, a legal connection can be established between the violation of the duty of care and the third party's criminal act. However, not all objects inherently carry the same level of risk. Although they do not pose a risk to the legal interests of others when used as intended and in a socially appropriate manner, they may become dangerous when used by inexperienced individuals. For such items, the duty of care cannot be extended to the same degree as for inherently dangerous objects. Imposing such a broad duty of care would unreasonably restrict the intended and socially appropriate use of these items[1683].

In summary, evaluations in this context consider the risks associated with the conduct (or system) in question, the likelihood of inexperienced individuals using it, and the ordinary flow of social life. In general, individuals of equal status are not obligated to monitor each other's behaviour. However, in certain hazardous activities, even colleagues of equal rank may be required to monitor one another. There exist duties of care specifically designed to enable individuals bound by them to address and mitigate the mistakes of others. While such duties are sometimes explicitly codified in positive legal norms, there are also unwritten sources of duty of care aimed at preventing harm and misconduct by others. In such circumstances, the perpetrator cannot invoke the principle of reliance to absolve themselves of liability[1684].

In light of the aforementioned debates, according to the principle of reliance, a manufacturer is entitled to assume that their products will be used correctly by consumers. However, this assumption depends on the manufacturer's obligation to provide clear and comprehensive information

[1682] Higher Regional Court of Stuttgart (OLG Stuttgart), judgment of 21.11.1996, Case No. 1 Ws 166/96, reported in NStZ 1997, p. 191.

[1683] PUPPE, § 5 Der Vertrauensgrundsatz in Strafrecht AT, 2023, p. 81 Rn. 1.

[1684] *Ibid*, p. 89 Rn. 22.

regarding potential risks associated with the use of the product[1685]. Furthermore, while manufacturers and autonomous systems may generally rely on humans to comply with established rules, they must also account for foreseeable errors, even in the absence of clear indications, due to the critical importance of safety and the capabilities of current technology. Such foreseeable errors include delayed reactions, such as those occurring in moments of shock, or sudden steering by human drivers. However, intentional[1686] or self-harming human actions should, as a general principle, not be taken into account. Conversely, errors that occur with statistical frequency should be incorporated into system programming. Empirical research is essential to determine which forms of human error are reasonably expected. A manufacturer who fails to account for such erroneous behaviour in the programming of their systems, at least as a potential scenario, acts negligently and may bear liability in the event of resulting damage[1687].

While this general observation provides an overview, further elaboration would help illuminate specific circumstances. For instance, in cases where a semi-autonomous vehicle detects a hazardous situation, it alerts the driver and requests them to take control, making it necessary for the driver to assume manual operation. According to one perspective, the driver's failure to assume control in such situations is a foreseeable circumstance from the manufacturer's standpoint. Consequently, the system should be designed to account for such a scenario, potentially by activating the hazard warning lights and bringing the vehicle to a stop through remote control mechanisms[1688]. In my view, while I agree that this situation is foreseeable from the manufacturer's perspective and that precautions should be taken accordingly, this approach risks unduly absolving the individual in the driver's seat from their responsibilities. It is essential to assess the matter based on the specific circumstances of each case. Furthermore, as highlighted within the frameworks of the *prohibition of regression* and *negligent*

[1685] ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 179.

[1686] According to one view, in situations such as traffic accidents, grossly negligent misconduct by the victim interrupts the chain of attribution (*Zurechnungszusammenhang*), whereas merely negligent misconduct does not. See: RENGIER, § 52. Das fahrlässige Begehungsdelikt in Strafrecht AT, 2019, p. 542 Rn. 56a.

[1687] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 453; HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 12; HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 154 f.

[1688] WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 230.

undertaking[1689], individuals operating such high-risk systems must possess the competence to take control of the vehicle when necessary to minimise potential dangers. To this end, it may be advisable for individuals intending to operate such systems to undergo basic training to ensure that they are adequately prepared for such situations. This would also ensure that the manufacturer's obligation to provide proper instructions is adequately addressed by the relevant parties.

A significant example in this context is a semi-autonomous driving accident that occurred in Switzerland in 2016, where a Tesla vehicle with its semi-autonomous autopilot features engaged[1690]. The driver, distracted by his phone, failed to pay attention to the road. As the vehicle approached a construction zone where the lanes had shifted, it failed to adjust its path, crashing directly into a signal trailer and a towing vehicle, causing significant property damage. The driver claimed that the autopilot malfunctioned and attempted to shift responsibility to Tesla. However, the court rejected this defence, highlighting the driver's primary obligation to maintain control and attention at all times while driving (it is worth noting that the absence of a legal provision akin to Section 1a of the StVG, introduced in Germany in 2017). The court further ruled that the driver's behaviour was not merely negligent but grossly negligent, given that the construction site was clearly visible, and the driver was evidently inattentive for at least 20 seconds before the collision. While the court's decision has been supported on the grounds that the autopilot technology at the time was not sufficiently advanced to be relied upon without question, attention has also been drawn to the challenges posed by the "control dilemma"[1691].

[1689] See: Chapter 4, Section C(3)(d): "Negligent Undertaking".

[1690] HOFSTETTER Johannes, "High-tech does not protect against punishment", 30.11.2017, https://www.bernerzeitung.ch/hightech-schuetzt-vor-strafe-nicht-399521855238. (accessed on 01.08.2025).

[1691] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 9 ff.
For another accident involving Tesla's autopilot, where the driver's hands were not on the steering wheel and the system had previously issued both visual and auditory warnings to place their hands back on the wheel, see: "Tesla in fatal California crash was on Autopilot", 31.03.2018, https://www.bbc.com/news/world-us-canada-43604440.
For example, as a good example of fulfilling duty of care, in the video shared by the user; the driver promptly intervenes and takes control due to their attentiveness, thereby preventing a potentially fatal manoeuvre by the autonomous driving system: https://x.com/thedooberhead/status/1869502131897782451?s=12. (accessed on 01.08.2025).

In this context, the 2007 decision of the Munich District Court (*Amtsgericht München*), although a civil law case, is noteworthy. In the case, a driver was held liable for damages when the parking assistance system they were using failed to signal due to a hollow space. The court emphasised that drivers cannot rely solely on such technology and must ensure safety through their own observation[1692].

Another issue concerns whether third parties can still be reasonably expected to act in full compliance with the rules in cases where, for example, a semi-autonomous vehicle experiences a minor malfunction. For instance, in a situation where the vehicle erroneously swerves into the wrong lane due to a minor malfunction, if other drivers on the road overreact, assuming that it is experiencing a serious malfunction, and this overreaction causes an accident or, as discussed above, if the driver assumes control despite the absence of a warning and an accident occurs as a result[1693]. Of course, the concept of error (*Irrtum*) could be applied in such cases. However, beyond this, it is necessary to evaluate the matter from the perspective of the principle of reliance.

In my view, particularly during the transitional period, as people become accustomed to the widespread adoption of AI-driven autonomous systems, machines should place less reliance on humans. This is because, currently, self-driving vehicles remain atypical for society. Therefore, it can be expected that people, upon noticing the absence of a driver in the vehicle, may react with confusion, leading them to make mistakes or behave in ways they would not normally. These machines, equipped with sensors capable of rapidly perceiving their surroundings, must account for and mitigate the potential for such atypical human behaviour. This necessity stems from the overarching duty to refrain from harm.

Furthermore, it can be argued that the principle of reliance 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, for instance, self-driving vehicles continuously perform risk as-

[1692] Local Court of Munich (AG München), decision of 19.07.2007, Case No. 275 C 15658/07, reported in NZV 2008, p. 35. THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 27 f.; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 287 f.

[1693] For a minor accident involving Google's semi-autonomous driving system and caused by a "misunderstanding", see: "Alex Davies, Google's Sel-Driving Car Caused Its First Crash", 29.02.2016, https://www.wired.com/2016/02/googles-self-driving-car-may-caused-first-crash/. (accessed on 01.08.2025).

sessments as part of their operation through their sensors and advanced computers, 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.

In this regard, the principle of reliance cannot be applied in exactly the same way to self-driving vehicles as it is to other road traffic participants. Instead, this principle should be considered solely in relation to the manufacturer's responsibility for certain foreseeable situations. For example, in the above-mentioned case of a pedestrian suddenly stepping onto the road to cross at a red light, the collision avoidance system of a self-driving vehicle must be developed to detect and respond to such scenarios, as the technology permits this level of precision. In situations where the vehicle perceives the pedestrian and manoeuvres accordingly, yet an accident still occurs, the applicability of permissible risk should be assessed based on the specific circumstances of the case. However, in line with the principle of reliance, the self-driving vehicle should not proceed at full speed without reducing its pace, relying solely on its right of way[1694]. 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 sensors scanning all other elements of the road environment.

(3) Should AI-Driven Autonomous Systems Rely on Each Other?

In the interaction between one autonomous system and another, the question arises as to whether they can rely on each other. In this context, in light of the foregoing explanations, what is ultimately at issue is whether the manufacturer can rely on whether the other systems will function correctly and reliably. For autonomous vehicles to operate safely in traffic, the coordination between road users that typically occurs in such settings is crucial[1695]. In particular, it is anticipated that self-driving vehicles will become widespread in the future and will communicate with each other as they navigate[1696]. In this regard, it may be possible to adapt a form of the principle of reliance for such networked systems. However, in this sce-

[1694] For a similar view, see also: WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 214.

[1695] KIRN/MÜLLER-HENGSTENBERG, Intelligente (Software-)Agenten, 2014, p. 231.

[1696] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 12.

nario, other autonomous systems that are not networked will be unable to integrate into this interaction. For these non-networked systems, the aforementioned explanations regarding the reliance of machines on humans remain applicable. Hence, they must be designed to take measures against foreseeable and expected misconduct. However, it is reasonable for them to operate under the assumption that entirely atypical situations beyond such design considerations will not occur.

E. Dilemma Challenges

1. Exploring the Origins of Moral Dilemmas

The introduction of AI-driven autonomous systems, in particular, self-driving vehicles into daily lives has reignited discussions surrounding the ancient moral dilemma. The 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 in German, English and Turkish legal literature. Numerous scholars have actively engaged in discussions suggesting that *Welzel*'s renowned "switchman" dilemma thought experiment[1697] has transitioned from theory to reality[1698]. All these discussions centre on addressing a fundamental question: how should a self-driving vehicle decide when faced with a dilemma?

This topic has inspired an extensive body of philosophical and legal literature, reflecting its enduring relevance and complexity. The ongoing ethical analyses by scholars on the matter demonstrate that determining the most correct choice remains challenging even today[1699]. Moral dilemmas have been the subject of various examples throughout history, with the question of what ethical choices should be made through numerous different variations. For instance, in the *Plank of Carneades*, two shipwrecked sailors face the moral quandary of deciding who gets to survive when only one can cling to a life-saving plank. Similarly, the famous *Trolley Problem* presents a moral dilemma of whether to pull a lever to redirect a trolley out

[1697] WELZEL, Zum Notstandsproblem, 1951, p. 51.

[1698] For example: SANDHERR, Strafrechtliche Fragen, 2019, p. 4.

[1699] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 146 f.; HILGENDORF, Dilemma-Probleme, 2018, p. 683 ff.

of control, sacrificing one person to save five[1700]. Another variation of this, in the *Fat Man*, one must decide whether to push a large person off a bridge to stop a runaway truck and save five others[1701]. These scenarios highlight the timeless nature of such moral dilemmas, challenging individuals to weigh competing ethical principles and responsibilities.

The increasing use of autonomous systems has led to frequent emphasis on the likelihood of encountering moral (or legal) dilemmas in real-life scenarios. Therefore, a legally valid conclusion to address the matter must be sought, regardless of the ethical deadlock on the matter. Because ethical principles and legal regulations may often diverge, reflecting significant differences in their nature and application[1702]. Hence, this study will examine the issue within the framework of existing criminal law mechanisms.

2. The Dilemma for Self-Driving Vehicles

a. How Does it Emerge?

In the context of AI-driven autonomous systems, whether the issue truly constitutes a dilemma akin to former moral dilemma examples is seldom debated. Instead, the focus often lies on the notion that a machine's decision in a dilemma scenario can be pre-programmed, making human biases and vulnerabilities in similar situations irrelevant, while raising the question of which decision would be morally and legally correct. Indeed, unlike humans, machines cannot make decisions influenced by emotions or exhibit tendencies to favour their loved ones, as they are inherently devoid of such biases[1703]. Similarly, a system can be programmed to prioritise saving or sacrificing pedestrians, animals, property, etc.

The dilemma for self-driving vehicles refers to scenarios where the vehicle, despite following traffic rules, is forced into an unavoidable accident and must choose to sacrifice one or more legal interests to save other(s). For instance, in a recently publicised incident, a vehicle driving in accor-

[1700] THOMSON, Killing, Letting Die, and The Trolley Problem, 1976; THOMSON, The Trolley Problem, 1985. Thomson refers to an earlier philosophical debate of Philippa Foot. See: FOOT Philippa, The Problem of Abortion and the Doctrine of Double Effect, 1967.

[1701] EDMONDS, Would You Kill the Fat Man, 2014, pp. 36-40.

[1702] ROBLES CARRILLO, Artificial Intelligence, 2020, p. 6.

[1703] ANDERSON/ANDERSON, Machine Ethics, 2007, p. 18.

dance with the rules, swerved left to avoid a pedestrian who suddenly fell onto the road, resulting in a collision with an oncoming car[1704]. Whether autopilot or human driver, for an external observer, the current scenario closely mirrors the very dilemmas debated in the literature concerning self-driving vehicles: a life has been saved at the expense of damage to the vehicles.

In the given incident, the pedestrian was saved thanks to the driver's quick reflexes; however, determining whether the driver consciously chose to risk property damage to save a life within milliseconds is nearly impossible. By contrast, when an accident becomes unavoidable, self-driving vehicles, owing to the processing power of the software, can rapidly evaluate all possible courses of action and select the option that minimises damage to the greatest extent possible[1705]. Therefore, the consensus is that dilemmas in autonomous driving are fundamentally different because, unlike human drivers who act reflexively without weighing *pros* and *cons* in unavoidable danger, autonomous systems are not constrained by such limitations, making previous scenarios and precedents largely irrelevant[1706]. Besides, saving passengers over pedestrians cannot be equated with the "human will to survive", which often exempts individuals from liability; whereas, normally, a human driver who endangers others to save themselves (albeit by committing an unlawful act) may not be held criminally liable since the law does not demand superhuman behaviour from people[1707].

In these new dilemmas, the vehicle's conduct is determined in the programming phase, long before the accident, rather than at the moment or immediately beforehand[1708]. Hence, there is no concept of "fate" or a "natural path" which the vehicle must follow[1709]. Thus, a pre-determined rational decision is implemented in practice. However, in a specific scenario, numerous uncertainties will arise simultaneously, making it nearly impossible to foresee all outcomes in advance. At the time of programming,

[1704] While the media widely portrayed this as the autopilot *heroically saving the pedestrian, in reality,* it was the human driver, through an instantaneous manoeuvre, saving the pedestrian. "Tesla autopilot heroically diverts collision to save pedestrian in Romania", 20.10.2024, https://en.as.com/videos/tesla-autopilot-heroically-diverts-collision-to-save-pedestrian-in-romania-v/. (accessed on 01.08.2025).

[1705] SCHUSTER, Das Dilemma-Problem, 2017, p. 100 f.

[1706] *Ibid*, p. 104.

[1707] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 574 f.

[1708] HEVELKE/NIDA-RÜMELIN, Selbstfahrende Autos, 2015, p. 10; BECK, Das Dilemma-Problem, 2017, p. 133.

[1709] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 11.

it is unclear whether a legal interest will be violated, which legal interests might be affected, or who might be involved specifically; only the general and abstract possibility of such violations can be anticipated. Consequently, no one holds a secure legal position during the programming phase[1710].

According to the prevailing opinion, such dilemmas in fact represent a subset of intentional crimes where the programmer's responsibility for the AI-driven system's decisions and subsequent conduct are questioned when the system causes an offense to avoid another legally prohibited outcome. Since the programmer must deliberately decide in advance how to program the vehicle, criminal liability for negligence is out of the question[1711]. Conversely, an alternative perspective[1712] will demonstrate the greater significance of liability for negligence.

Although self-driving vehicles are anticipated to cause fewer accidents overall compared to human drivers, their widespread use will inevitably result in harm to certain legal interests. In dilemmas, determining which legal interests should be prioritised and which should be sacrificed is a moral and legal challenge. For instance, should the vehicle prioritise the safety of its passengers or pedestrians? The young or the elderly? Humans or animals? The educated or the less educated? More lives over fewer lives? Moreover, these distinctions are not always straightforward or directly identifiable, adding further complexity to the issue.

One of the factors contributing to the contemporary popularity of the topic is the *Massachusetts Institute of Technology* (MIT)'s online experiment called *Moral Machine*[1713]. Although the results were not based on strict scientific criteria, they provide a rough insight into global trends. Setting legally valid conclusions aside, the experiment highlights that ethical preferences vary significantly across different regions and demographics. By 2018, the experiment had gathered approximately 40 million decisions, in ten languages, from millions of participants across 233 countries and territories. The prominent findings include a global preference for sparing more lives (quantity); prioritising humans over animals and showing a local tendency to protect younger individuals[1714].

Setting ethical choices aside, the legal decision to be adopted involves numerous factors to be carefully considered; such as the hierarchy of values

[1710] *Ibid*; FELDLE, Notstandsalgorithmen, 2018, p. 63.

[1711] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 180.

[1712] See: Chapter 4, Section E(4): "Evaluation: An Alternative Approach".

[1713] https://www.moralmachine.net. (accessed on 01.08.2025).

[1714] AWAD, et al., The Moral Machine Experiment, 2018.

and whether quantitative and qualitative comparisons are feasible. When one legal interest is sacrificed to save another, it is crucial to determine which legal principle applies (for instance, whether the conditions of necessity or conflict of obligations are applicable) on a case-by-case basis. Furthermore, additional specific issues arise, including the probability of injury (harm), the focus on self-protection, and the distinction between action and omission[1715].

In any case, manufacturers introducing self-driving vehicles to the market are obligated to equip their vehicles with collision avoidance systems and comprehensive protocols for dilemma-like situations to address all foreseeable situations. Failing to develop coping strategies for these scenarios may constitute a breach of their duty of care due to design defects and the absence of required safety standards may potentially lead to criminal liability[1716].

b. The Balancing of Interests

(1) Comparison of Values

As will be analysed in detail below, under German law, pursuant to Section 34 of the German Criminal Code (StGB), the application of necessity as a justification requires that the protected legal interest be one of life, limb, liberty, honour, property or another legally recognised interest. Furthermore, the protected interest must substantially outweigh the interest that has been infringed upon to meet the proportionality requirement necessary for justification. On the other hand, under Section 35 of StGB, the application of necessity as an excuse requires that the protected legal interests be limited to life, limb, or liberty. In this context, it is essential to evaluate whether the conditions of these legal constructs are met through a detailed examination, alongside an analysis of which interests and values may be at stake in the dilemmas encountered by self-driving vehicles. Hence, it is crucial to determine whether these interests and values can be prioritised over others, and whether a protected interest substantially outweighs the one being infringed upon.

[1715] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 160 ff.

[1716] FELDLE, Notstandsalgorithmen, 2018, p. 197, 252.

Self-driving vehicles utilising AI can categorise their environment and are programmed to make decisions that align with predefined safety priorities. In such dilemmas, if the protected interest does not significantly outweigh the impaired one, such conduct is deemed unlawful[1717]. To determine which ones outweigh others, a hierarchy of legal interests must first be established[1718]. Determining which value holds greater importance may not always be straightforward. This assessment could be guided by examining the penalties prescribed for criminal offences under the special provisions of penal codes, as these reflect the legal interests they aim to protect[1719].

The legal interests associated with self-driving vehicles centre primarily on the protection of human life, holding supreme importance; whether it concerns passengers, pedestrians, or other individuals. Closely linked to this is the safeguarding of physical integrity, a critical legal interest that is particularly vulnerable to violation in the event of traffic accidents. Additionally, the protection of property emerges as another major legal interest; encompassing damage to vehicles, infrastructure, and other material assets. In dilemmas, while pursuing the necessity of protecting certain endangered legal interests, the infringement of others becomes inevitable. Therefore, the values being infringed upon and those being protected must first be identified[1720].

Although it may be challenging to make a choice in certain scenarios, the almost unanimous opinion is that life holds the utmost value which should not be questioned[1721]. It is generally accepted that physical integrity follows life in importance, with material values ranked thereafter. However, adopting an abstract categorical approach may be difficult. For instance, would a few bruises be considered an acceptable trade-off for saving tens of thousands of Euros[1722]?

Although such scenarios are unlikely to arise in self-driving vehicles, other autonomous systems may encounter dilemmas where state interests conflict with other legal values such as human life. In light of past debates in literature, it is generally asserted that life should be prioritised above all else in such cases[1723]. Although the abstract principle that human life can

[1717] ENGLÄNDER, Das selbstfahrende, 2016, p. 380.

[1718] FELDLE, Notstandsalgorithmen, 2018, p. 105.

[1719] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 238 Rn. 236.

[1720] RENGIER, § 19. Rechtfertigender Notstand in Strafrecht AT, 2019, p. 183 Rn. 26.

[1721] FELDLE, Notstandsalgorithmen, 2018, p. 187.

[1722] *Ibid*, p. 116.

[1723] *Ibid*, p. 111.

never be equated with property is logical, it has been argued that exceptions may arise; *e.g.* in the context of a fire, a document of critical importance may take precedence over an individual's life if it holds significant implications for saving many others[1724]. Nonetheless, it is crucial to adopt a cautious approach to such discussions.

Under German law, the principle of solidarity operates on the rationale that the protection of substantially outweighing legal interests justifies the sacrifice of lower-level interests. This principle reflects a balance between individual and collective responsibilities. However, the sanctity of life is regarded as inviolable and remains exempt from this expectation, underscoring its supreme legal and moral value[1725]. In a legal system grounded in human rights and dignity, solidarity does not demand self-sacrifice[1726].

In this regard, apart from the evaluations for necessity in criminal law, the German Road Traffic Act (StVG) (Section 1e(2)(2)) addresses dilemmas with provisions designed to prevent and minimise damage. It specifies that, in cases of unavoidable alternative harm to different legal interests, the significance of these interests must be considered, prioritising the protection of human life above everything. Furthermore, it explicitly prohibits any further weighing of human lives based on personal characteristics, such as age or gender. Thus, the legislation aims to implement ethical guidelines for autonomous driving. However, this remains a highly complex matter, raising unresolved ethical, legal, and technical challenges. While it is generally agreed that human life and physical integrity take precedence over property in such scenarios and that human lives are not to be weighed against each other based on qualitative characteristics, the technical feasibility of these guidelines remains uncertain. Besides, more complex issues, such as deciding between multiple lives, are still far from being resolved[1727].

When addressing such dilemma questions, the Ethics Commission on Automated and Connected Driving, established by the German Federal Ministry of Transport and Digital Infrastructure, emphasised that general programming should aim to minimise the number of personal injuries. It further concluded that sacrificing one person's life to save others would not be lawful[1728].

[1724] ÖZEN, Öğreti ve Uygulama, 2023, p. 679.

[1725] FELDLE, Notstandsalgorithmen, 2018, p. 110.

[1726] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 805.

[1727] HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 448.

[1728] For detailed discussions, see: Ethik-Kommission Automatisiertes und Vernetztes Fahren, Bericht der Ethik-Kommission Automatisiertes und Vernetztes Fahren, In a case where a self-driving vehicle is faced with a situation in which it must choose between causing injury to an individual or colliding with a barrier, thereby causing damage to property, it is appropriate to conclude that the less significant right should be sacrificed in accordance with the principles of conflicting interests[1729]. However, in certain instances, comparing the values at stake may prove exceedingly complex, leading to choices where every possible outcome corresponds to a tragic scenario and constitutes a breach of the law[1730].

While most dilemmas typically involve conflicts between different types of legal interests, rare instances may present life versus life conflicts[1731]. In such scenarios, both quantitative debates, such as sacrificing one person to save several others as in the classical trolley problem; and qualitative discussions, such as prioritising the life of a younger person over that of an older individual, fall within this scope. Although the initial reaction might suggest that saving a greater number of people in an unavoidable situation is preferable, according to the established view the sacrifice of an innocent person cannot be justified on the basis that it would result in saving another or even a greater number of lives[1732].

In *Kantian* philosophy, every individual is regarded as possessing inherent dignity, an absolute value distinct from a price, and therefore cannot be subjected to valuation or comparative assessment in terms of worth[1733]. Reflecting this principle, German criminal law, deeply rooted in *Kantian* deontological ethics, deems it morally impermissible to actively cause harm, even to save others[1734]. Intentionally killing an innocent person is never justified. The inherent value of each life is regarded to be maximum, and multiple lives are not considered more valuable than a single life[1735]. Con-

Bundesministerium für Verkehr und digitale Infrastruktur, June 2017, https://bmdv.bund.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile. (accessed on 01.08.2025). HILGENDORF, Autonome Systeme, 2018, p. 107; HILGENDORF, Dilemma-Probleme, 2018, p. 682.

[1729] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 146 f.; EREM, Ümanist Doktrin, 1971, p. 38

[1730] HILGENDORF, Dilemma-Probleme, 2018, p. 692.

[1731] BECK, Das Dilemma-Problem, 2017, p. 119.

[1732] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 173 f.

[1733] HILGENDORF, Recht und autonome Maschinen, 2015, p. 24.

[1734] See: NEUMANN, Recht und Moral, 2021, p. 13.

[1735] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 151; RENGIER, § 19. Rechtfertigender Notstand in Strafrecht AT, 2019, p. 184 Rn. 32; SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 10; ZIESCHANG, Strafrecht AT, 2023, p. 76 Rn. 259.

sequently, in such dilemma situations, Section 34 of StGB requires refraining from action (e.g., not switching tracks) to avoid "playing fate", as the law permits interference with another's interests only when the protected interest significantly outweighs the one being compromised. However, this provision for necessity does not apply in such cases because all lives are considered equal in value. This approach contrasts with the consequentialist perspective prevalent in Anglo-American law, which may justify actions that lead to the best overall outcome[1736].

The absolute protection of life is a cornerstone of German legal tradition and has been debated in various contexts over decades. For instance, Section 14(3) of the Aviation Security Act (*Luftsicherheitsgesetz*), which authorised the interception and destruction of a passenger plane being used to kill others, was declared unconstitutional[1737]. Through this, the German Constitutional Court has upheld *Kant*'s assertion that no human being may be reduced to a mere means, even in the pursuit of a noble end[1738]. This aligns with Article 1 of the German Constitution (*Grundgesetz*), which stipulates that human dignity shall be inviolable, and Article 19(2), which stipulates that the essence of a fundamental right may not be infringed under any circumstances. Hence, unlike other fundamental rights, there is no exception and any restriction will be unlawful[1739]. To concede that there is no alternative to avoid danger other than killing an innocent third party ultimately equates to acknowledging the existence of a "right to kill"[1740].

The absolute prohibition against quantifying human life is also reflected in the criminal laws of other countries, including Belgium, Switzerland and Austria. In some legal systems, however, the standard of "equivalence of legal interests" is deemed sufficient, rather than "substantially outweigh". In contrast, U.S. law predominantly argues in favour of justifying the killing of individuals to save many, reflecting a more consequentialist approach[1741]. In Turkish law, however, it is sufficient for there to be a proportionality between the gravity of the danger, the subject matter and the means used, according to Article 25(2) of TPC.

[1736] JOERDEN, Zum Einsatz, 2017, p. 81.
[1737] FELDLE, Delicate Decisions, 2017, p. 200.
[1738] JOERDEN, Zum Einsatz, 2017, p. 93.
[1739] HILGENDORF, Dilemma-Probleme, 2018, p. 685.
[1740] EREM, Ümanist Doktrin, 1971, p. 42.
[1741] FELDLE, Notstandsalgorithmen, 2018, pp. 215-217.

(2) Assessment of the Utilitarian Approach to Dilemmas

Could the sacrifice of a single individual ever be justified to save multiple lives, such as five people? What if it were one hundred? What, then, should be done if a hijacked plane is heading towards a nuclear power plant situated near a city inhabited by millions of people[1742]? Would it be legally approved to sacrifice a terminally ill patient to save an entire train of passengers? Or to justify the sacrifice of a fleeing bank robber, who unintentionally created such a dilemma, in order to save a bus full of students returning from school? How about instances in which a shared danger threatens a group of people and the sacrifice of some could guarantee the survival of others? While these questions present significant challenges, the prevailing legal and moral perspective indisputably rejects such behaviour, with evaluations of the latter scenario being treated as a distinct consideration.

In dilemmas, while adjusting crash optimisation, ethical guidelines, although individual *if-then* formula cannot be constructed for all alternative scenarios in the world, can technically be embedded into self-driving vehicles' decision-making algorithms. However, the fundamental challenge lies in determining how these ethical or legal norms should be applied -whether based on deontological strict rules or focused on the outcomes of decisions in a consequentialist manner[1743]. For instance, if all other factors remain constant, a key question is whether ethical principles should guide decisions, such as prioritising collisions (causing injury) with those violating rules (such as colliding with individuals crossing at a red light)[1744]. Additionally, there is a debate over whether societal (or even religious) values should influence the interpretation of these norms[1745]. Nonetheless, even in such scenarios, sacrificing an individual cannot be permitted to undermine the fundamental protection of human dignity enshrined in Article 1 of the German Basic Law (*Grundgesetz*)[1746].

In the classical examples provided in literature, such as the mountain climber cutting the rope to save themselves, or the Plank of Carneades, most of society may morally approve the climber cutting the rope, in the context of balancing competing interests. However, every human life holds

[1742] JOERDEN, Zum Einsatz, 2017, p. 93.
[1743] GERDES/THORNTON, Implementable Ethics, 2016, p. 88 ff.
[1744] LIN, Why Ethics Matters, 2016, p. 73.
[1745] OTTO, Pflichtenkollision, 1965, p. 49.
[1746] HILGENDORF, Dilemma-Probleme, 2018, p. 697 f.

the same intrinsic value, regardless of how much time to live remains for an individual or the certainty of their death. Consequently, the climber cannot invoke necessity as a valid defence[1747]. Moreover, according to the prevailing opinion, prioritising one life over another is impermissible, and factors such as age, gender or ethnic background cannot serve as valid considerations in such decisions[1748]. Besides, in evaluations involving quantitative calculations, sacrificing one person to save five, as in the switchman case, may be morally applauded by society. However, if the example is slightly altered, for instance where a doctor sacrifices a completely healthy individual to save five patients, would not receive the same approval[1749].

While rejecting the inclusion of human life as part of the equation is a principled stance rooted in respect for human dignity and fundamental values, it may not resolve the practical challenges that arise. In a dilemma, unavoidable danger necessitates a decision. Although sacrificing one individual to save many cannot be justified; still, there remains a moral and legal obligation to minimise the number of fatalities[1750].

Approaching the issue analytically under the general principle of minimising harm inevitably leads to quantitative calculations, if not qualitative ones, and leads to an examination of consequentialist approaches. Unlike in Germany, offsetting human lives is not considered a taboo in Anglo-American legal traditions[1751]. In this context, attention is drawn to the possibility of utilising the results of an experiment (like Moral Machine), albeit not scientific, in which millions of people worldwide participated and which reflects the diverse values of societies, could represent a more reasonable approach for autonomous driving by aiming to achieve the greatest benefit and satisfaction for society through a utilitarian framework[1752].

In classical utilitarianism, moral actions are evaluated based on their outcomes rather than their inherent moral or legal meanings[1753], assigning

[1747] ZIESCHANG, Strafrecht AT, 2023, p. 77, 104 Rn. 262, 372.

[1748] Ranking individuals based on qualitative characteristics is not only contrary to human dignity but also, as recent German history demonstrates, such approaches can lead to extremely dangerous consequences. See: HILGENDORF, Dilemma-Probleme, 2018, p. 695; HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 805.

[1749] FELDLE, Notstandsalgorithmen, 2018, p. 233.

[1750] HILGENDORF, Recht und autonome Maschinen, 2015, p. 26.

[1751] FELDLE, Notstandsalgorithmen, 2018, p. 249.

[1752] SEUFERT, Wer fährt, 2022, p. 326.

[1753] HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 806.

numerical weights to each potential result to maximise overall utility[1754]. In the context of autonomous driving, this framework advocates prioritising actions that minimise harm, particularly in unavoidable crash scenarios, by saving the greatest number of lives[1755]. However, determining which choice brings more "utility" requires calculating what constitutes a good or bad outcome, as *Bentham*'s perspective is fundamentally a form of moral arithmetic[1756].

Adopting a utilitarian approach in autonomous driving entails mathematically optimising outcomes, potentially making it computationally feasible for algorithmic decision-making. However, this approach faces significant challenges in quantifying harm and valuing human life, which raises significant ethical and legal concerns, including risks of discrimination and conflicts with principles of equality and the right to life[1757]. Moreover, applying classical utilitarianism could lead to harm for third parties not directly involved in the dilemma, as they too may be sacrificed for the greater good[1758]. Furthermore, although it may provide a convenient mathematical framework for quantifiable calculations, qualitative situations cannot be calculated, therefore will always remain uncertain. Additionally, it would inevitably result in the consistent sacrifice of particular groups for utilitarian purposes, which is equivalent to intentional killing or injuring; therefore, is not legally justifiable[1759].

In an instance of three children suddenly running onto the road during lawful driving; where doing nothing would result in all three dying; swerving left would kill one and swerving right would kill two; with all risks being entirely equal, none of these choices can be legally justified[1760]. However, the concept of a gradation in injustice becomes relevant here. Both ethically and legally, swerving left would be the necessary course of action to at least save the life of a child. Although literature includes views

[1754] Utilitarianism, a form of consequentialism, emerged primarily to promote the broadest possible distribution of welfare. Although it played a significant role in the 19th and 20th centuries, particularly in combating slavery and shaping parliamentary democracy, it has traditionally been regarded with contempt in Germany. See: HILGENDORF, Dilemma-Probleme, 2018, p. 686 f.

[1755] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, p. 329.

[1756] ANDERSON/ANDERSON, Machine Ethics, 2007, p. 18.

[1757] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, p. 329 f.

[1758] HILGENDORF, Dilemma-Probleme, 2018, p. 687.

[1759] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, p. 330.

[1760] For the example, see: HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 156.

suggesting that no intervention (going straight without swerving) should be made in such cases, inaction itself would also constitute a decision in the context of autonomous driving. Assessing such scenarios as a matter of faith and remaining inactive would be inappropriate[1761]. It should be noted that, in this example, all three children are subject to the same danger, and intervention is made to choose the option that quantitatively results in fewer casualties. In other words, intervention saves at least two lives, as otherwise, all would certainly die. Allowing all to die would indeed be an absurd choice. In such scenarios where individuals face a shared danger, the number of potential victims must be considered when making decisions. Killing is not legally permissible; however, deliberately choosing to kill three children instead of one, when two could have been saved, contradicts the principle of choosing the lesser evil. This matter will be further discussed below under supra-legal necessity.

(3) Proximity of Danger, Impact of Predictable Decisions and Random Generator

The question of whether option A or B should be chosen in dilemma scenarios are based on the assumption that the desired outcome will be definitively achieved by selecting an option. In other words, it is based on the abstract premise that choosing option A will certainly result in the loss of B but the gain of A, and *vice versa*. However, such clear-cut scenarios are exceedingly rare in real-life situations. Therefore, as will be elaborated below, it can be argued that classical moral-dilemma-like scenarios are unlikely to arise in the context of self-driving vehicles. Instead, the duty of care for mitigating the risks and the scope of permissible risk should be made the point of assessment.

In any case, collision avoidance systems must aim to reduce risks which encompass both the probability and severity of danger. However, reducing the risk for one individual may create one for another. For instance, if a child suddenly runs onto the road while a vehicle is driving lawfully and the brakes are applied forcefully to avoid hitting the child, a motorcyclist approaching from behind may collide with the vehicle and is likely to suffer fatal consequences. In such rapidly developing situations, where harm is unavoidable, these systems can effectively and rapidly calculate all variables

[1761] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 156.

to implement the most optimal choice. Nevertheless, the balance of risks between the parties becomes an issue once again, depending on which legal interest is chosen for protection[1762]. In this context, it could be argued that reducing the risk faced by those most likely to suffer the greatest harm, such as death, would be an appropriate approach[1763].

For an unoccupied self-driving vehicle, it is sensible to prioritise sacrificing itself in an unavoidable collision, accepting property damage, to safeguard higher level legal interests; an expectation not commonly placed on human decision-making[1764]. However, when passengers are present inside the vehicle, the risks in such dilemmas will almost never be equal: While some individuals are inside the vehicle, others may be on bicycle, and some crossing the street with their dogs, or in other situations[1765]. This complicates the assessment of the status and likelihood of the infringement of legal interests that may occur.

A self-driving vehicle does not necessarily need to place its passengers in a disadvantaged position compared to others[1766]. Yet, in a collision scenario, rather than running over a pedestrian who typically faces the highest risk of severe harm, it is legally and morally preferable to program the vehicle to crash into barriers, thereby exposing its passengers, protected by seatbelts, to less severe risks in comparison to those faced by the pedestrian. However, this gives rise to another issue: it is likely that the owners and manufacturers of self-driving vehicles may be reluctant to embrace this approach, particularly if such pre-programming is publicly known. Consequently, they may opt for programming which prioritises the protection of their passengers, contrary to the principles discussed here[1767]. This approach, particularly when these systems are widely implemented, would systematically disadvantage certain individuals and groups while expecting sacrifices from them[1768]. This raises the question of whether clear

[1762] OTTO, § 8 Pflichtbegrenzende Tatbestände in Grundkurs Strafrecht, 2004, p. 149 Rn. 202 ff.; FELDLE, Notstandsalgorithmen, 2018, p. 161.

[1763] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, p. 331 f.

[1764] HU, Robot Criminals, 2019, pp. 500-501.

[1765] HILGENDORF, Dilemma-Probleme, 2018, p. 698.

[1766] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 170.

[1767] HILGENDORF, Dilemma-Probleme, 2018, p. 698; HILGENDORF, Recht und autonome Maschinen, 2015, p. 27; MALGIERI/PASQUALE, Licensing High-Risk AI, 2024, p. 5.

[1768] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, pp. 333.
Prioritising their owners may lead to dangerous outcomes. For instance, in a dilemma between two pedestrians, the system might calculate that colliding with a

legal rules should be established to govern such scenarios in order to ensure a fair approach.

As previously stated, the outputs of AI are *ex ante* unpredictable and *ex post* difficult to explain. However, in attempting to minimise risks in dilemma situations by prioritising certain legal interests, their outputs become more foreseeable. Should it become possible to anticipate how these vehicles will decide or manoeuvre under specific circumstances, they could be exploited or manipulated for adversarial purposes[1769]. For instance, if a self-driving vehicle must choose between two motorcyclists -one wearing a helmet and the other not- it may prioritise colliding with the helmeted rider based on the lower likelihood of severe harm. Nevertheless, such programming would disadvantage the helmeted motorcyclists and, more broadly, those who follow safety rules[1770]. Moreover, this example is not exclusive to motorcyclists or individual circumstances; generalising this approach could result in the perpetual disadvantage of certain groups. Furthermore, individuals who recognise this general strategy may exploit it. Suddenly in a traffic dominated by self-driving vehicles, not wearing a helmet could ironically become a strategy that offers greater protection to the rider[1771]. To take the example further, travelling alone in a car could become a disadvantage, as a self-driving vehicles may have a stronger incentive to save the lives of a greater number of people in a dilemma situation[1772]. Similarly, warnings such as "Caution: Baby on Board" might become more widespread, even when there is no baby in the car. Moreover, this phenomenon could extend beyond traffic and influence other areas where AI-driven autonomous systems are employed. For instance, individuals awaiting organ transplants might deliberately neglect their health, inflict self-harm, or take other measures to manipulate the AI's evaluation system in order to appear more urgent or in greater need.

To prevent such abuse, the decisions made by autonomous systems should incorporate a degree of uncertainty. One proposed approach in-

wealthy individual poses higher compensation risks and instead target a less affluent person, such as a poor student. Such scenarios risk systematically disadvantaging certain groups. For the example, see: HU, Robot Criminals, 2019, p. 504 f.

[1769] OSÓRIO/PINTO, Information, 2019, p. 40.

[1770] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 162; GOODALL, Ethical Decision, 2014, p. 62; OKUYUCU ERGÜN, Machina Sapiens, 2023, p. 745; LIN, Why Ethics Matters, 2016, p. 73; OSÓRIO/PINTO, Information, 2019, p. 41.

[1771] SCHÄFFNER, Caught Up in Ethical Dilemmas, 2018, p. 330.

[1772] HEVELKE/NIDA-RÜMELIN, Selbstfahrende Autos, 2015, p. 14.

volves two potential methods: introducing noise into the decision-making process (internal uncertainty) or keeping the specifics of the system's decision-making and evaluation processes confidential (external uncertainty)[1773]. However, while adding noise introduces vagueness into system functioning, it also reduces decision-making quality. On the other hand, creating external uncertainty by making it difficult for third parties to observe and understand how the system operates might be effective in the short term; but maintaining the confidentiality of the process over the long term presents significant challenges[1774].

One perspective argues that, since the subjects of the specific incident are not known at the time of programming, minimising the number of victims and reducing the risk of collision will undoubtedly align with the interests of everyone. Here, the most rational choice should be made based on the information available at the time the programming. It is also stated that, as there is no information available during the programming regarding the parties involved in potential future accidents; programmers operate under conditions analogous to *John Rawls' veil of ignorance*. Therefore, it is proposed that they should develop programming that adheres as closely as possible to this moral principle[1775]. However, this view has been criticised for being legally unconvincing when making an ethical choice and ultimately leading to utilitarian consequences[1776].

The concept of making decisions randomly has been proposed as a solution to mitigate the risk of exploitation stemming from the predictability of a self-driving vehicle's decisions while also addressing the inherent complexities and deadlocks of dilemma situations. Accordingly, in the absence of viable outcomes from other rational solutions, it is questioned whether self-driving vehicles should address dilemmas by making entirely random decisions, thereby distributing risk equally. In real-life scenarios, individuals confronted with the possibility of a sudden accident, often make instinctive decisions; resulting in harm to one party and the survival of another, without conducting a detailed evaluation of all relevant factors. Such actions are generally regarded as lawful by society. Consequently, it is

[1773] OSÓRIO/PINTO, Information, 2019, p. 41.

[1774] *Ibid*, 2019, p. 43 ff.

[1775] HEVELKE/NIDA-RÜMELIN, Selbstfahrende Autos, 2015, p. 11 f. For a review, see: ENGLÄNDER, Das selbstfahrende, 2016, p. 378.

[1776] FELDLE, Notstandsalgorithmen, 2018, p. 191.

argued that a similar approach could be deemed acceptable for autonomous vehicles[1777].

However, the use of a random generator has been subject to considerable criticism. While it may resemble the spontaneous and incalculable reaction of a human driver, this does not make it a better solution[1778]. Developers of autonomous vehicles are not compelled to act randomly in situations of complete uncertainty, as they have access to extensive data and contextual factors that could enable the generation of potentially better solutions in some scenarios[1779]. Moreover, relying on randomness could allow manufacturers to evade liability under criminal law by hiding behind the element of chance[1780]. Fundamentally, not all individuals are subjected to equal risk in such situations. In a dilemma, one party's likelihood of harm may far exceed that of others, and there may also be other immeasurable considerations at play. Thus, random decision-making is not only unacceptable but also potentially unlawful[1781]. This situation can be compared to organ allocation through lotteries for transplant recipients. Even these lotteries are not entirely random, as systems often allocate more chances to patients with greater need[1782]. It has been argued that random generators can only be used in rare circumstances where a typical conflict of obligations situation arises, as in such cases, any choice could be justified, and absolute equality of opportunity could be effectively guaranteed[1783].

3. Legal Frameworks Applicable to Dilemma Situations

Under this section, the main legal constructs applicable to dilemmas will be examined, including necessity as a justification, necessity as exculpation, supra-legal excusable necessity and the conflict of obligations. Rather than examining all aspects of these legal frameworks, the focus will be on the dimensions relevant to dilemma scenarios. In addition to these, the consent

[1777] FELDLE, Delicate Decisions, 2017, p. 202 f.; FELDLE, Notstandsalgorithmen, 2018, p. 202 f.

[1778] SCHUSTER, Das Dilemma-Problem, 2017, p. 110.

[1779] HILGENDORF, Recht und autonome Maschinen, 2015, p. 22; FELDLE, Delicate Decisions, 2017, p. 202 f.

[1780] JOERDEN, Zum Einsatz, 2017, p. 88.

[1781] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 453 Rn. 51; FELDLE, Notstandsalgorithmen, 2018, p. 212 f.

[1782] *Ibid*, p. 207.

[1783] *Ibid*, p. 212 f.

of the individual involved is proposed as another applicable legal construct, suggesting that a person might choose to sacrifice themselves to save the lives of multiple others. However, such instances are unlikely to be generalised, and the legal system cannot expect anyone to sacrifice themselves. Consequently, this perspective has gained little support[1784].

a. Analysis under German Law

The concept of necessity is primarily categorised in two forms in accordance with *Differenzierungstheorie*: necessity as justification and necessity as exculpation[1785]. Determining the legal nature of necessity is not merely a matter of theoretical classification but is significant due to its impact on the resulting legal outcomes. For instance, in cases of necessity as justification, legitimate self-defence cannot be invoked against an individual acting out of necessity, whereas it can be invoked in cases of necessity as exculpation[1786]. Moreover, under justification, there is no liability for damages, even under civil law, as the act is considered lawful within the entire legal system. In contrast, under exculpation, only criminal liability is excluded, while civil liability remains intact[1787].

(1) Necessity as Justification (StGB Section 34)

According to Section 34 of the German Criminal Code (StGB), an act committed to avert an imminent danger to life, limb, or other legal interests is lawful if the protected interest substantially outweighs the one infringed, based on a balancing of the conflicting interests and the degree of danger. The *ratio legis* of this provision lies in the principle of solidarity, which justifies the violation of a legal interest by requiring individuals to tolerate the infringement of lower-value personal interests when confronted with a substantially greater legal interest[1788].

When balancing conflicting interests, all legitimate interests affected by the conflict must be taken into account. This includes factors such as

[1784] For the evaluation, see: *Ibid*, p. 58 f.

[1785] ZIESCHANG, Strafrecht AT, 2023, p. 104 Rn. 371.

[1786] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 382.

[1787] ÖZEN, Öğreti ve Uygulama, 2023, p. 766.

[1788] FELDLE, Notstandsalgorithmen, 2018, p. 59.

the actual extent of the damage to be expected, the nature, intensity and proximity of the danger, the potential losses, the relative importance of the legal rights involved, specific duties (*e.g.*, those of police officers, soldiers, or guarantors), the purpose pursued by the actor, the irreplaceability of potential damages, and the likelihood of successful intervention[1789].

The software of the system should be programmed to prioritise the option that causes the least harm in a dilemma[1790]. However, the core issue lies in establishing a hierarchy of harm and injuries that aligns with legal principles and ethical expectations[1791]. The violation of a legal interest to avert danger is justified under necessity only if the affected legal interests substantially outweigh those that are interfered with. It has already been discussed that when legal interests of differing types and degrees come into conflict, resolving the dilemma becomes straightforward if one substantially outweighs the other[1792]. For example, a driver may justify hitting a parked bicycle to save their own life[1793].

For self-driving vehicles, a key issue in dilemmas is their programming to strictly follow traffic rules. In some cases, avoiding an accident may require breaking a minor rule, such as driving onto the pavement. The software must permit such conduct to prevent greater harm. In other words, the aim is to avoid a more severe outcome by permitting a lesser rule violation. The legal challenge is determining in advance which violations are less severe, given the unpredictability of real-world scenarios[1794]. For example, in a dilemma, should lightly touching a pedestrian resulting in extremely minor injury be considered preferable to the vehicle being completely destroyed and incurring significant financial loss?[1795] Moreover, real-life scenarios do not always mirror the classic moral dilemma of sacrificing one person to save three. For instance, even if the system prioritises saving three individuals, its calculations might show a 40% chance of hitting one individual if it swerves left, versus a 5% chance of hitting two individuals if

[1789] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 469; FREUND, § 3 Fehlende Rechtfertigung, 2009, p. 96 Rn. 65.

[1790] HILGENDORF, Automated Driving and the Law, 2017, p. 189.
Engländer compares the pre-programming of specific commands for dilemmas to *Offendicula*, such as automated self-defence systems (e.g., high-voltage fences). See: ENGLÄNDER, Das selbstfahrende, 2016, p. 376.

[1791] HILGENDORF, Automated Driving and the Law, 2017, p. 189.

[1792] See: Chapter 4, Section E(2)(b)(1): "Comparison of Values".

[1793] FELDLE, Delicate Decisions, 2017, p. 197.

[1794] *Ibid*, p. 198.

[1795] *Ibid*.

it swerves right. What should be done in such an instance? In my view, the focus should shift away from classic dilemma scenarios towards approaches that minimise risk and align with the concept of permissible risk, as this represents a more practical and preferable approach in line with real-life circumstances.

Conflicts between two legal interests of equal value can also present challenges. In real-time situations, it is often difficult to determine the hierarchical significance of two abstractly defined interests in practical terms. For instance, there is no doubt that a minor injury is "substantially outweighed" by a severe injury. However, when it comes to damage involving two property interests, should the financial cost of one outweighing the other be the determining factor? What if one of them carries significant sentimental value? The most critical theoretical debate centres on whether sacrificing one person to save one or more other person(s) constitutes a "substantial outweighing" of interests. This brings into focus the divergence between utilitarian and deontological perspectives, that have been discussed above. The core issue lies in determining how an autonomous system should be programmed to address such dilemmas[1796]. In an unavoidable situation of this kind, while killing one person instead of two may seem preferable at first glance; the legal basis for reaching such a conclusion remains unresolved[1797].

As detailed above, in contrast to the utilitarian-leaning Anglo-American legal system, the German legal system regards every life as holding maximum value, rejecting any quantitative comparison of the right to life. Therefore, sacrificing one life cannot be deemed to substantially outweigh even the saving of tens of lives. Consequently, the prevailing and nearly unanimous opinion is that even in an emergency, it is unlawful to kill one person to save two others and necessity as justification does not apply[1798]. While this principle is clear, there is also an ethical and legal obligation to minimise harm and the number of fatalities[1799].

[1796] FELDLE, Delicate Decisions, 2017, p. 199.

[1797] HILGENDORF, Autonome Systeme, 2018, p. 108.

[1798] For the consistent jurisprudence on the impermissibility of sacrificing one life to save others, and its determination that such actions constitute a clear violation of human dignity, see: Federal Court of Justice (BGH), judgment of 28.11.1952, Case No. 4 StR 23/50, reported in NJW 1953, p. 514.
HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 96, Rn. 83; HILGENDORF, Automated Driving and the Law, 2017, p. 190; FELDLE, Delicate Decisions, 2017, p. 200.

[1799] HILGENDORF, Automated Driving and the Law, 2017, p. 190.

Despite the prevailing opinion, an alternative view holds that a person whose life is endangered is justified under Section 34 of the StGB to severely injure or, in extreme cases, even kill the person causing the danger[1800]. Furthermore, in the case of a hijacked plane[1801], shooting down the aircraft can be legally justified. This rationale prioritises saving people on the ground and contends that, while the passengers are innocent, they bear some degree of responsibility for the ongoing danger by virtue of being part of the flight, unlike the individuals on the ground who are entirely uninvolved[1802]. To frame a question for scholarly discussion, one might ask whether this perspective could be extended to scenarios involving self-driving vehicles, where passengers, though mostly passive, benefit from delegating transportation to an autonomous system. In other words, can such passengers be regarded as the source of the danger and thus given lower priority compared to uninvolved third parties? In my view, the answer to this question should be negative, although their liability should be separately discussed for delegating a task to autonomous systems.

Autonomous systems driven by AI may also play a role in decision-making across various areas, such as organ transplantation or blood donation scenarios, where dilemmas may also arise. One example of the numerous dilemmas that may arise in this context is the case of a critically injured patient with a rare blood type. If an individual with a matching blood type arrives at the hospital but refuses to donate, the question emerges whether it would be lawful to forcibly take blood from them (through a harmless medical procedure) to save the patient's life. The prevailing view holds that this would not be legally justified under Section 34 of the StGB, because even if the protected interest outweighs the impaired one, solidarity cannot be mandated and assistance in such cases remains an act of moral freedom[1803].

[1800] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 239 f. Rn. 247.

[1801] See: Chapter 4, Section E(2)(b)(2): "Assessment of the Utilitarian Approach to Dilemmas".

[1802] GROPP/SINN, § 5 Rechtswidrigkeit in Strafrecht AT, 2020, p. 241 Rn. 251.
See for discussions: LADIGES, Die notstandbedingte, 2008, p. 131 f., 140.

[1803] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 364; FRISTER, 17. Kapitel - Strafrecht Allgemeiner Teil, 2020, p. 244 Rn. 15.
For a discussion, whether human dignity may take precedence over the interest in preserving life, see: ROXIN/GRECO, § 16. Der rechtfertigende Notstand in Strafrecht AT, 2020, p. 860 Rn. 48 ff.

(2) Necessity as Exculpation (StGB Section 35)

According to Section 35 of the German Criminal Code (StGB), a person who commits an unlawful act to avert imminent danger to their own life, limb, or liberty -or that of a relative or close person- acts without guilt. In this provision, "body" refers to physical integrity, and "freedom" pertains specifically to the physical freedom of movement, rather than the broader concept of general freedom of action[1804].

The key distinction of necessity as exculpation under Section 35 of the StGB from necessity as justification under Section 34, lies in the limitation of the types of legal interests protected. Unlike justification, exculpation does not require the protected legal interest to substantially outweigh the one infringed, aligning with its focus on culpability. Another significant difference is that the relevant legal interests must pertain to the individual themselves, a relative or a close person. Another distinction is that, unlike necessity as justification, in cases of necessity as excuse, the individual's actions may be unlawful, making self-defence against them admissible[1805].

The reason necessity as justification exempts an offender from punishment is not their subjective reaction to the psychological situation they face; rather, it is based on the objective reality that, in such circumstances, anyone would be compelled to harm another's legal interest. In contrast, necessity as excuse applies when an individual is under exceptional psychological duress that makes lawful behaviour unreasonable to expect; thereby diminishing the wrongfulness and culpability of their illegal act[1806]. This psychological state can be explained by moral coercion or the instinct of self-preservation[1807]. Accordingly, under moral coercion, individuals forced to make split-second decisions in moments of danger are not influenced by the fear of punishment, as their actions are not the result of deliberate, calculated choices. Consequently, such behaviour lacks social dangerousness[1808]. To speak of pressure on the perpetrator, they, their relative or

[1804] RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 246 Rn. 5; KINDHÄUSER/HILGENDORF, § 35 Entschuldigender Notstand - Strafgesetzbuch, 2022, p. 335 Rn. 3.

[1805] RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 244 Rn. 1.

[1806] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 683; RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 244 Rn. 1.

[1807] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 378.

[1808] EREM, Ümanist Doktrin, 1971, p. 39; ÖZEN, Öğreti ve Uygulama, 2023, p. 758. For an evaluation of the same view, see: ENGLÄNDER, Das selbstfahrende, 2016, p. 381.

a close person must be in imminent danger[1809]. Moreover, the law does not require acting in a state of panic as a condition; otherwise, composed individuals would face punishment while those who panic would remain unpunished[1810]. For these reasons, the scope of necessity as an exculpatory defence is accurately limited. Unlike self-defence, this doctrine often involves harm to uninvolved innocent third parties[1811].

In a real collision scenario, even an experienced human driver typically lacks the time and ability to calculate the least harmful course of action, often relying on reflexes to choose the most reasonable option in that moment. Machines, however, do not face this limitation; with powerful processing capabilities, they can scan the entire environment within milliseconds, process data in line with current conditions, and calculate the probability of a crash. Therefore, they should be equipped with crash optimisation strategies[1812].

The rationale that the condition of psychological pressure on the perpetrator will make the application of this provision extremely challenging in dilemmas involving AI-driven autonomous systems. This is because the programmer is not in an acute mental crisis or tragic decision-making situation at the time of the offence; therefore, *ex ante* reliance on an excuse is not possible[1813]. In contrast, the decisions in question are pre-programmed and based on rational choices[1814]. Furthermore, it has been argued that the danger could have been avoided if they had refrained from programming the autopilot in the first place[1815].

Another challenge in applying necessity as exculpation to dilemmas involving AI-driven autonomous systems is the condition that the danger must be directed at the offender themselves or at a relative or close person. However, it is evident from the discussed dilemma examples that neither the manufacturer nor the programmers, whose criminal liability may be assessed, are relatives or closely connected to those at risk, such as passengers, drivers, or third parties on the road facing imminent danger.

[1809] MERAKLI, Ceza Hukukunda Kusur, 2017, p. 383 fn. 117 & 118.

[1810] FELDLE, Notstandsalgorithmen, 2018, p. 67.

[1811] EREM, Ümanist Doktrin, 1971, pp. 40-41.

[1812] HILGENDORF, Recht und autonome Maschinen, 2015, p. 21; LIN, Why Ethics Matters, 2016, p. 75, 81.

[1813] SCHUSTER, Das Dilemma-Problem, 2017, p. 106; SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 10; JOERDEN, Zum Einsatz, 2017, p. 87; ENGLÄNDER, Das selbstfahrende, 2016, p. 381.; SEUFERT, Wer fährt, 2022, p. 327.

[1814] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 452 Rn. 49.

[1815] JOERDEN, Zur strafrechtlichen, 2020, p. 296 f.

Therefore, Section 35 of the StGB would not apply[1816]. This inference can be extended to other similar examples involving AI-driven autonomous systems, where the application of necessity as excuse under Section 35 would be extremely challenging.

One perspective on this matter suggests that instead of discussing whether manufacturers or programmers can invoke necessity as excuse for their pre-programming decisions; the focus should shift to the individual activating the vehicle and evaluating their proximity to those at risk. Accordingly, this person would be aware of and accept the manufacturer's pre-programmed decisions; fully cognisant of the circumstances under which specific choices will be implemented. Ultimately, this individual would be the one who ultimately sets the actual risk[1817]. This perspective definitely approaches the issue from a reasonable standpoint. However, it could be argued that, in real-life scenarios, it is unlikely that individuals would fully comprehend all the options for which the AI system has been trained. Rather, it involves accepting the potential risks of using such a system with only an approximate understanding of them. Moreover, another issue arises in invoking necessity as excuse in such cases: the condition that the individual must not have caused the danger. Causing the danger should not be interpreted according to the condition theory; otherwise, its scope of application would become overly broad and even permissible behaviours would fall within this scope, significantly narrowing the application of Section 35 of the StGB[1818].

In conclusion, it could be argued that neither of the necessity provisions under Sections 34 and 35 of the StGB can generally be applied to dilemmas involving AI-driven autonomous systems, particularly in cases involving self-driving vehicles where the killing of another is at issue. However, while necessity as justification may not apply due to the "substantially outweigh" condition, it is argued that, in extremely exceptional cases, such as the Plank of Carneades, the killing of another person may be excused, for example, when a shipwrecked individual pushes another off a rescue plank that can support only one person[1819]. This issue will be further discussed below under *supra-legal excusable necessity*.

[1816] BECK, Das Dilemma-Problem, 2017, p. 133; SEUFERT, Wer fährt, 2022, p. 327, SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 10.

[1817] FELDLE, Notstandsalgorithmen, 2018, p. 96 f.

[1818] For the discussion on causing the danger, see: RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 248 Rn. 18.

[1819] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 689.

(3) Supra-Legal Excusable Necessity

In situations requiring a choice between the lives of multiple individuals, neither necessity as justification nor necessity as exculpation appear to provide a legal solution to the dilemmas involving AI-driven autonomous systems, particularly self-driving in road-traffic. For instance, in *Welzel*'s switchman example, where a railway switchman must decide whether to redirect a train to save many lives at the cost of three[1820], the individual cannot rely on Sections 34 or 35 of the StGB. This is because sacrificing a life is impermissible, and the people saved are not their close relatives.

It has been doctrinally and almost unanimously accepted that life cannot be weighed against life. However, one might consider a scenario involving a hijacked airplane carrying innocent passengers and being directed toward a residential area. While shooting down the plane, thereby sacrificing those aboard, would be unconstitutional; what practical measures should be taken in such a case? Can the potential deaths of tens of thousands of uninvolved and innocent residents simply be disregarded? Moreover, what should be done if the hijacked plane is heading toward a nuclear power plant located near a densely populated city of millions[1821]?

The situation becomes particularly complex when every possible choice appears to be legally impermissible. However, in cases where all potential victims are exposed to the same danger and at least a quantitative decision can be made, the dilemma becomes more nuanced. For instance, in a scenario where three children jumped onto a road and steering right would result in the deaths of two children, steering left would cause the death of one child, and taking no action would lead to the deaths of all three[1822]. Determining the appropriate programming is challenging in such a dilemma. Steering left and sacrificing one life to save two violates the prohibition against quantifying and weighing human lives. Conversely, failing to save the maximum number of lives could contravene the principle that human life holds the highest value. It is inherently contradictory to classify human life as the “highest value” while at the same time arguing that the loss of one, two, or three lives is of no significance. Whereas the death of one person is tragic, the loss of two or all three lives is undoubtedly worse.

[1820] WELZEL, Zum Notstandsproblem, 1951, p. 51.

[1821] For the example, see: JOERDEN, Zum Einsatz, 2017, p. 93.

[1822] For the example, see: HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 15 f.

Therefore, in situations where all potential victims face the same danger, the priority must be to save as many lives as possible. Hence, in such dilemmas, choosing the lesser evil is the most pragmatic solution. However, this approach inherently means that the absolute prohibition against quantifying and weighing human life cannot be maintained[1823].

As another example, a human driver, through no fault of their own, enters a road that ends abruptly without any warning signs. At the end of the road, there are 20 children playing on one side and a single individual on the other, and there is no time to stop the vehicle. If the driver swerves at the last moment to collide with the single individual instead of the children, neither necessity as justification nor exculpation applies since the driver is not personally at risk, nor are the children their close relatives, and multiple lives do not substantially outweigh one. Thus, it is proposed to apply *supra-legal excusable necessity* in such exceptionally rare cases, with strict consideration of specific conditions[1824].

Section 35 of the StGB is often inapplicable due to its restrictive provision limiting its scope to the protection of oneself, close ones or relatives[1825]. According to the prevailing opinion, the requirements and restrictions of Section 35 of the StGB generally apply to the supra-legal excusable necessity[1826]. However, with respect to the necessity as excuse, a threat to life alone is fundamentally sufficient. This threat must place the perpetrator in a state of mental conflict comparable to that experienced when their own life or the lives of their close relatives are at risk; yet this does not necessarily need to involve only close relatives[1827]. In such situations, the solution is simpler when all potential victims are exposed to the same danger and would die regardless of intervention. By contrast, the more challenging scenario involves making a quantitative assessment, where individuals who were not previously in danger are sacrificed to save the majority[1828].

[1823] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, pp. 15-16.

[1824] HILGENDORF, Autonomes Fahren im Dilemma, 2017, pp. 160-161; ENGLÄNDER, Das selbstfahrende, 2016, p. 368 ff.

[1825] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 128 Rn. 58

[1826] RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 253 Rn. 43.

[1827] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 711 ff.; KINDHÄUSER/ZIMMERMANN, § 24 Entschuldigender Notstand - Strafrecht AT, 2024, p. 214 Rn. 18

[1828] According to Rengier, the prevailing opinion also supports the extension of supra-legal necessity to include uninvolved parties. See: RENGIER, § 26. Entschuldigender Notstand in Strafrecht AT, 2019, p. 253 Rn. 44 f.

Some views in literature assert that supra-legal necessity applies only in instances where individuals are already exposed to danger and would ordinarily die in the absence of intervention. It does not extend to uninvolved third parties, *e.g.* passengers on a hijacked plane already facing mortal risk[1829]. Therefore, collision avoidance systems should not be programmed to manoeuvre in a manner that sacrifices individuals not previously at risk in order to save a greater number of lives[1830].

Situations where the entire group faces a life-threatening danger and only some can be sacrificed to save the rest are not approved by case law, particularly in light of the stipulations set forth in Article 1 of the German *Grundgesetz*. However, according to the widespread opinion in literature, a distinction must be made between asymmetrical and symmetrical danger groups. In asymmetrical danger groups, certain individuals who are already doomed to die are sacrificed to save the rest of the group. For example, a ship captain may isolate a specific section of a sinking ship, leaving those within it to perish while ensuring the survival of the rest. By contrast, in symmetrical danger groups, any specific subset of individuals from the group must be sacrificed to save the rest; for instance, on an overcrowded lifeboat where certain individuals must be sacrificed for the survival of the others; otherwise, everyone would perish. In the asymmetrical danger group scenario, since those sacrificed are already destined to die, strict adherence to an absolute prohibition on killing would counterproductively undermine the principles of protection and lesser evil: the killing of each innocent person is legally wrong; but the number of innocent victims must be kept as low as possible[1831]. As a result, sacrificing these individuals is more widely accepted in literature. However, in the symmetrical danger group scenario, where everyone has an equal chance of survival, the issue becomes far more controversial. It could be argued that it seems irrational for the law to demand the death of the entire group when some could have been saved[1832].

[1829] HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 128 Rn. 58; RENGIER, § 19. Rechtfertigender Notstand in Strafrecht AT, 2019, p. 185 Rn. 35.

[1830] HILGENDORF, Automatisiertes Fahren als Herausforderung, 2019, p. 15 f.

[1831] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 174; HILGENDORF, Moderne Technik, 2015, p. 109.
The principle of lesser evil applies not only to vehicle collision dilemmas but also to situations like breaking into a house to survive in freezing conditions. See: HILGENDORF, Dilemma-Probleme, 2018, p. 690.

[1832] KINDHÄUSER/ZIMMERMANN, § 17 Rechtfertigender Notstand - Strafrecht AT, 2024, p. 174 f. Rn. 30 ff.; HILGENDORF, Dilemma-Probleme, 2018, p. 702.

In light of the explanations regarding supra-legal necessity, in dilemmas involving self-driving vehicles, where a collision is imminent and the vehicle has no third alternative or the possibility to brake; it can be argued that the vehicle should be programmed to minimise damage by choosing the lesser evil[1833]. In such cases, the application of supra-legal necessity may be a relevant consideration. However, such scenarios might involve placing individuals who were not initially at risk into a position of danger, which cannot be justified. Even in instances of risk redistribution, the sacrifice of individuals who were not previously endangered would remain unlawful, as no one can be expected to sacrifice their life for the benefit of others[1834].

The application of supra-legal necessity has been subject to criticism from various perspectives in legal literature. It has been argued that the supra-legal necessity would usually fail due to its very narrow conditions[1835]. According to one view, such an excuse, which should already be limited to exceptional circumstances, risks leading to an unbounded expansion due to the concept of supra-legal necessity, and therefore should not be applied[1836]. Another view holds that its application undermines the condition explicitly stipulated in law that an excuse should only be granted if the person in danger is either themself or someone close to them[1837].

Another criticism is that, even in situations where a group of individuals face the same danger, sacrificing those who are destined to die to save others is still unacceptable. This is because even one second of their lives is not inherently less valuable than the potentially longer lives of those who might be saved[1838]. And finally, the same criticism for necessity as exculpation has been put forward, as supra-legal necessity is inapplicable to self-driving vehicles because the programmer, as the decision-maker,

[1833] HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 30; HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 805.

[1834] HILGENDORF, Dilemma-Probleme, 2018, p. 692; HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 805; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 714 ff.
For the view that sacrificing uninvolved individuals to save more lives cannot either be considered under permissible risk, see: JOERDEN, Zum Einsatz, 2017, p. 87 f

[1835] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 478.

[1836] ZIESCHANG, Strafrecht AT, 2023, p. 109 f. Rn. 386 ff.

[1837] JOERDEN, Zum Einsatz, 2017, p. 77 f.

[1838] WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 476.

operates under no immense pressure and makes decisions rationally and intentionally[1839].

(4) Conflict of Obligations

In criminal law, it is widely recognised that the grounds for excuse or justification are not *numerus clausus*. There may be other grounds that exist beyond the legally defined ones of self-defence, necessity as justification and exculpation. These include supra-legal necessity and justifying conflict of obligations[1840].

In a genuine conflict of obligations, there are multiple binding obligations, and it becomes necessary to fulfil one while acting contrary to the demands of other(s)[1841]. In such cases, the principle of *ultra posse nemo obligatur* applies; exceptionally permitting the disregard of one obligation for the sake of fulfilling the competing one[1842].

A conflict of obligations arises when an individual faces multiple obligations; but can fulfil only one at the expense of others. In other words, the individual can fulfil both obligations; but cannot fulfil them simultaneously. Here, a value assessment is conducted to determine the appropriate course of action. The weight of the competing duties is assessed according to the principles governing the standard of Section 34 of the StGB, taking into account the value of the endangered interests and the respective probability of harm. In cases where obligations are of equal importance, Section 34 cannot be invoked, as no obligation significantly outweighs the other;

[1839] FELDLE, Notstandsalgorithmen, 2018, p. 101.

[1840] SATZGER, StR Die rechtfertigende Pflichtenkollision, 2010, p. 753.
Under the German Criminal Code (StGB), three primary perspectives have been advanced regarding the legal status of an individual's actions under conflict of obligations: unlawful; unlawful but excused; or unlawful and culpable, yet subject to a justification excluding punishment. See: OTTO, Pflichtenkollision, 1965, pp. 66-70.
In his 1965 work, Otto examined the applicability of the necessity provisions in the 1962 draft of the German Criminal Code (StGB) to cases of conflict of obligations but did not offer a definitive solution to the issue. See: OTTO, Pflichtenkollision, 1965, p. 114. The current StGB closely mirrors the 1962 draft's necessity provisions but introduces exceptions for cases involving self-created danger or special legal relationships, allowing punishment mitigation.

[1841] OTTO, Pflichtenkollision, 1965, p. 48.

[1842] KINDHÄUSER/HILGENDORF, § 34 Rechtfertigender Notstand - Strafgesetzbuch, 2022, p. 333 Rn. 57.

however, unlike Section 34, the individuals are expected to prioritise the slightly higher interest[1843]. If one duty substantially outweighs the other -such as saving a life versus protecting property- necessity as justification can still be invoked. Then again, in cases where the obligations are of equal value, the individual is free to choose which duty to fulfil, and disregarding the other is legally justified; rather than merely excused[1844]. However, in cases where two non-equivalent duties conflict, it is not legitimate to fulfil the lesser duty while disregarding the one of greater value[1845].

Obligations may conflict in different ways; such as an active obligation conflicting with an obligation to refrain, or two obligations to act conflicting with each other. In such situations, the slightly higher obligation takes precedence. The prioritisation is not determined solely by the value of the legal interests linked to the obligations but also by an assessment of the overall interests at stake, the perpetrator's intended objective, and widely accepted societal values[1846].

An example of a conflict of obligations is when a lifeguard must choose between saving one of two drowning individuals. In such a case, the actor is free to decide, and as long as the legal system does not prescribe the correct course of action, their conduct cannot be subsequently disapproved[1847]. As another example, an obligation to act may conflict with an obligation to refrain, as in the case of a doctor needing to breach patient confidentiality in order to warn others of a potential risk of infection[1848].

In cases where the conflicting obligations are of equal value in terms of the legal interests involved and all other relevant circumstances, a distinction must be made regarding the type. When an obligation to act conflicts with one to refrain, the general principle is to prioritise refraining from action; meaning that the individual should remain passive. A situation in which a single ventilator is available and is already being used for a patient

[1843] *Ibid*, Rn. 58 ff.

[1844] RÖNNAU, Vor §§ 32 ff in LK, 2020, p. 118, Rn. 124; KINDHÄUSER/ZIMMERMANN, § 18 Rechtfertigende Pflichtenkollision - Strafrecht AT, 2024, p. 182 f. Rn. 3 ff.
Roxin/Greco considers such conflict of obligations as supra-legal justification. See: ROXIN/GRECO, § 16. Der rechtfertigende Notstand in Strafrecht AT, 2020, p. 889 Rn. 122.

[1845] RÖNNAU, Vor §§ 32 ff in LK, 2020, p. 116, Rn. 122; KINDHÄUSER/ZIMMERMANN, § 18 Rechtfertigende Pflichtenkollision - Strafrecht AT, 2024, p. 182 Rn. 5.

[1846] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 365 f.

[1847] SCHUSTER, Das Dilemma-Problem, 2017, p. 108 f.

[1848] JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 366.

can be given as example. Removing the ventilator from the first patient to save another, resulting in the death of the initial patient, would not be approved by law. If, however, removing the ventilator would only slightly injure the first patient, while saving the second patient's life, it can be argued that necessity as an excuse may be invoked. Similarly, when two obligations to act or two obligations to refrain conflict -for instance, if a doctor must choose between saving one of two equally critical patients arriving at the hospital simultaneously- saving one at the expense of the other's life is excusable[1849]. Moreover, in an intensive care unit, the termination of an ongoing treatment to commence the saving of another person's life cannot be justified through a conflict of obligations, as it involves two equally valuable interests: the right to life[1850].

A classic example frequently discussed in literature involves a scenario where a fire simultaneously breaks out in both wings of a hospital, raising the question of whether the firefighter should prioritise saving a larger group of individuals in one wing or those in the other wing with fewer people[1851]. One perspective posits that the correct course of action would be to rescue the larger group. However, this raises the question of whether, under German law, failing to save the smaller group would constitute a failure of duty[1852]. In contrast, another view emphasises that human lives cannot be reduced to mere numbers, asserting that each life holds maximum value. Accordingly, both choices are considered equally valid and legal[1853]. Yet, this scenario differs fundamentally from the case of shooting down a hijacked plane[1854] where it became an instrument[1855] and individuals are

[1849] KINDHÄUSER/ZIMMERMANN, § 18 Rechtfertigende Pflichtenkollision - Strafrecht AT, 2024, p. 183 Rn. 7. According to the authors, the minority opinion asserts that this constitutes an excuse, indicating that when the norm addressee selects one option, their behaviour is deemed justifiable.
For the discussion regarding justification and exculpation in such instances, see: JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, pp. 366-368.

[1850] RÖNNAU, Vor §§ 32 ff in LK, 2020, p. 117, Rn. 123.

[1851] MERKEL, § 14 Abs. 3 Luftsicherheitsgesetz, 2007, p. 380.

[1852] *Ibid.*

[1853] FELDLE, Delicate Decisions, 2017, pp. 200-201.

[1854] *Ibid*, p. 200.

[1855] The instrumentalization of a person, or their treatment as a "mere object" occurs when they are killed solely because they pose a source of danger, as in the case of a child manipulated by terrorists into becoming an unwitting threat. See: MERKEL, § 14 Abs. 3 Luftsicherheitsgesetz, 2007, p. 382.

actively forfeited and killed to save a larger number of uninvolved potential victims[1856].

Finally, it is essential to address the applicability of the conflict of obligations to the dilemmas encountered by self-driving vehicles. These discussions primarily focus on whether making an active choice (e.g., swerving the steering wheel) constitutes an act of commission or whether refraining from intervention qualifies as an omission[1857]. For example, in the scenario where three children suddenly run onto a road, actively intervening could kill one or two children, whereas taking no action might result in all three being killed[1858]. While the traditional approach favours non-intervention, this approach does not apply to self-driving vehicles, as even inaction of the vehicle stems from pre-programming[1859]. In such cases, due to the deadlock, the legislator's intervention may be considered; yet it may be plausible to accept the absence of criminal liability if at least one of the superior or equally significant obligations is prioritised[1860].

According to one view, in dilemmas involving self-driving vehicles, two active obligations do not come into conflict. Therefore, the recognised principles for justifying conflicts between equivalent obligations cannot serve as a basis for granting the obligation-bearer the right to choose between fulfilling one or another equally significant obligation. In this context, it cannot be asserted that there is a conflict between the active obligation not to kill the single child on the left or the two children on the right and the passive obligation not to kill all three. Consequently, no genuine choice exists in such a scenario. Moreover, doctrinal issues surrounding the conflict of entirely passive obligations exist, which makes their applicability in this context highly questionable[1861].

Based on another view, in dilemmas involving self-driving vehicles, an active obligation conflicts with an obligation to refrain. However, such conflicts can only be resolved through the application of Section 34 of the StGB, which permits the infringement of previously uninvolved legal interests if the protected interest significantly outweighs the infringed one (but it does not in present cases)[1862].

[1856] MERKEL, § 14 Abs. 3 Luftsicherheitsgesetz, 2007, p. 381.

[1857] FELDLE, Notstandsalgorithmen, 2018, p. 72 ff.

[1858] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 156.

[1859] BECK, Das Dilemma-Problem, 2017, p. 133.

[1860] *Ibid*, p. 134.

[1861] JOERDEN, Zum Einsatz, 2017, p. 90 f.

[1862] FELDLE, Notstandsalgorithmen, 2018, p. 102.

Another standpoint based on Swiss law argues that, in such dilemmas, where there is no higher-value distinction between two lives at stake, necessity as a justification is inapplicable. However, the criminal liability of the programmer could potentially be excluded under the concept of justifying conflict of obligations. This would apply if the situation involves two equally valuable legal interests and the programmer is unable to design the software in a manner that ensures the preservation of both lives in the event of an accident[1863].

b. Analysis under Turkish Law

In Turkish law, there is only one provision potentially relevant to the topic: *necessity* stipulated under Article 25(2) of the Turkish Penal Code (TPC). According to this provision, *no penalty shall be imposed on the perpetrator for acts committed with the necessity to save oneself or another person from a grave and certain danger, which is directed against one's own or another's right, which is not caused knowingly and which cannot be protected in any other way, and provided that there is a proportion between the severity of the danger and the subject and the means used*[1864].

The first notable aspect of the provision in Turkish law is that, unlike necessity as a justification in German law, the law only requires proportionality rather than the substantial outweighing of one legal interest over another. In other words, the provision only refers to proportionality between the severity of the danger, the subject and the means employed. It does not address a balance between the legal interests sacrificed and those preserved. Furthermore, unlike both necessity provisions in German law, Turkish law imposes no restrictions regarding the type of rights involved. Additionally, the perpetrator may act out of necessity to save any third party, without the requirement that the individual be a relative or a person with a close relationship to the perpetrator.

The legal nature of this provision in Turkish law is not explicitly defined in the statute, and it has been a subject of debate in legal literature. In brief,

[1863] MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 180.

[1864] The translation was made by the author. For another English translation, see: Council of Europe, European Commission for Democracy through Law (Venice Commission), Penal Code of Turkey, Opinion No. 831/2015, CDL-REF(2016)011, 15 February 2016, https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2016)011-e. (accessed on 01.08.2025).

it exhibits characteristics of both justification and an exculpatory excuse, and its scope of application is determined accordingly.

The following observations can be made regarding whether the necessity provision in Turkish law constitutes a justification or an excuse: the fact that the mere endangerment of any legally protected right is sufficient, and that the danger may threaten either the perpetrator's rights or those of another, are characteristics of a justification. On the other hand, the absence of a requirement for a substantial value difference between the protected and sacrificed rights, as well as the condition that the perpetrator must not have knowingly caused the danger, are features of an exculpatory excuse[1865].

In addition to these, the phrase "no penalty shall be imposed" within the provision, in conjunction with the fact that Article 25 is stipulated under Part 2 of the TPC titled "Grounds Excluding or Diminishing Criminal Liability" does not assist in clarifying its legal nature. Furthermore, although not binding, an explanatory memorandum on the provision explicitly describes it as a ground for exculpation. Additionally, Article 223(3) of the Turkish Criminal Procedure Code specifies that, in offences committed under a state of necessity, the perpetrator is considered to lack culpability. In light of the aforementioned facts, it has been posited that the legal nature of the provision in Turkish law cannot be considered as justification[1866]. It is further argued that, as in German law, having two separate provisions for necessity would be more appropriate in Turkish law[1867].

[1865] MERAKLI, Ceza Hukukunda Kusur, 2017, p. 384; ÖZEN, Öğreti ve Uygulama, 2023, p. 764 f.
According to one view, the equivalent of necessity as exculpation in German law is *compelling reason** under Turkish law. In this case, the perpetrator acts out of desperation and under severe psychological pressure, making it unreasonable to expect compliance with the norm. The necessity provision in the TPC can serve as a basis for both justification and *compelling reason*. See: ZAFER, Ceza Hukuku, 2021, p. 461 f.
* This term, rather than *force majeure* has been adopted. Because the author here conceptualises the concept as forces that compel the perpetrator to engage in a particular course of conduct in an irresistible and unavoidable manner.

[1866] MERAKLI, Ceza Hukukunda Kusur, 2017, p. 382 ff.; ÖZGENÇ, Türk Ceza Hukuku, 2019, pp. 435-438; AKBULUT, Ceza Hukuku, 2022, p. 663 f.
Nonetheless, it is argued that in order to apply necessity in Turkish law as an excuse, the provision must specify its scope by clarifying the legal interests and individuals to which it applies, thereby narrowing its scope. See: MERAKLI, Ceza Hukukunda Kusur, 2017, p. 470.

[1867] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 385.

One perspective in Turkish legal literature emphasises the importance of interpreting necessity in a manner that ensures its broad application[1868]. However, this approach raises concerns, particularly when the protected and sacrificed legal values are of equal importance[1869]. The current provision, for instance, equates the right to life of an innocent uninvolved third party with that of the individual whose life is intended to be protected in dilemmas[1870]. Nevertheless, in cases where one value significantly outweighs the other, it can be treated as a ground for justification; whereas in cases where the values are equal, it may be regarded as a ground for excuse[1871]. On the other hand, due to the exceptional nature of necessity and for providing grounds for the breach of a right, it has been emphasised that the protected right must either be equal to or more significant than the sacrificed right in Turkish law[1872].

The aforementioned assessments under German law are similarly relevant in the context of Turkish law. However, there are notable divergences in the conclusions reached for dilemma situations in accordance with the provisions of the TPC. Remarkably, since there are no strict "substantially outweighing" conditions regarding the balance between the infringed and protected legal interests, the preference in a dilemma can lean towards saving a greater number of lives. Furthermore, the prohibition against comparing lives, which is a firm principle in German law, is not as prevalent in Turkish legal dogmatics[1873]. Additionally, although one view -rightly, from a theoretical perspective- argues that necessity for the benefit of others should be limited to specific individuals, such as relatives, or to situations where the protected interest outweighs the sacrificed one[1874]; this interpre-

[1868] ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 439.

[1869] ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 384; MERAKLI, Ceza Hukukunda Kusur, 2017, p. 384 fn. 121.
For the assessment that rather than determining which value is absolutely superior, the focus can be placed on which value is, in the ordinary course of life, deemed more worthy of protection, see: HAKERI, Ceza Hukuku, 2022, p. 394.

[1870] MERAKLI, Ceza Hukukunda Kusur, 2017, p. 387.

[1871] ÖZEN, Öğreti ve Uygulama, 2023, p. 760.
For the discussion that if the protected right significantly outweighs the sacrificed one, it should be considered a justification; or if they are equal or slightly outweighs, it should be treated as an excuse, see: ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 380, 384.

[1872] AKBULUT, Ceza Hukuku, 2022, p. 669.

[1873] However, in my view, the inviolability of human life and the significance of human dignity necessitate the strict application of this principle in Turkish law as well.

[1874] TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 175.

tation would not be feasible in practice given the explicit wording of the law. Therefore, the legal conclusions that were deemed inapplicable in dilemmas under German law may find application in Turkish law, allowing the perpetrator to rely on the defence of necessity.

Additionally, discussions on the conflict of obligations are similarly addressed in Turkish legal literature, particularly by scholars who engage with German legal doctrines. Accordingly, the conflict of obligations is a justification ground similar to the state of necessity, though not explicitly codified in the law[1875]. The analyses and examples provided in this context are closely parallel to those made under German law, but should be examined with due regard to the specific provisions of Turkish law. Hence, such dilemmas can also be evaluated under the conflict of obligations[1876].

Another aspect that requires examination under Turkish law is the condition that danger must not have been caused knowingly. To illustrate, in situations where a self-driving vehicle, operating lawfully, encounters a dilemma, could it be argued that the person behind the machine (particularly the programmer) knowingly caused the danger and therefore cannot invoke the defence of necessity? According to the prevalent opinion in literature, a reasonable benchmark should be applied and the danger should be interpreted as having been caused directly[1877]. It is generally accepted that the term "knowingly" in the provision encompasses only intent and conscious negligence[1878] (*bewusste Fahrlässigkeit*). Thus, a programmer who causes a dangerous situation through simple (unconscious) negligence (*unbewusste Fahrlässigkeit*) may invoke the necessity defence. However, in cases of erroneous programming that could be classified as conscious negligence, the programmer would not be able to rely on the necessity defence.

[1875] ZAFER, Ceza Hukuku, 2021, p. 415.

[1876] ÖZEN, Öğreti ve Uygulama, 2023, p. 677.
If no conclusion of superiority of a legal interest can be reached after all evaluations, fulfilling one of the obligations should be deemed excusable. ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 342 f.
For an analysis differentiating the conflict of obligations and the conflict of interests and, additionally, the scenario where a caretaker can only save one of two babies during a flood, can be assessed as a conflict of obligations as an excuse, see: ÖZEN, Öğreti ve Uygulama, 2023, p. 678.

[1877] EREM, Ümanist Doktrin, 1971, p. 45.

[1878] TOROSLU/TOROSLU, Ceza Hukuku, 2019, p. 176; HAKERI, Ceza Hukuku, 2022, p. 391; KOCA/ÜZÜLMEZ, Türk Ceza Hukuku, 2019, p. 349; ÖZEN, Öğreti ve Uygulama, 2023, p. 761, 772; ÖZBEK/DOĞAN/BACAKSIZ, Türk Ceza Hukuku, 2019, p. 389.

Finally, it should be emphasised that legal frameworks, shaped by numerous factors including the moral codes of different countries, may vary significantly. Accordingly, the software of self-driving vehicles developed and manufactured in one country must be adapted to ensure compatibility with the legal systems of other jurisdictions where they will be utilised[1879].

4. Evaluation: An Alternative Approach

This section of the study discussed the longstanding ethical dilemmas and their legal implications, with particular emphasis on the expectation that such dilemmas will become increasingly prevalent with the widespread adoption of self-driving vehicles. In this context, the moral and legal approaches that could be adopted when weighing conflicting values have been discussed, and the legal frameworks that may be applicable have been analysed. When a decision must be made between equivalent interests, such as the lives of two individuals; it is concluded that, despite the alternative perspectives presented in literature, German law does not provide a definitive solution through legal constructs such as necessity or conflict of obligations.

Turning back to the instance of the three children suddenly running onto a road during lawful driving (where doing nothing would result in all three dying, swerving left would kill one, and swerving right would kill two[1880]); assuming all risks are entirely equal and the outcome is certain through the appropriate manoeuvre, the programmer faces four potential options. These are: refraining from programming any specific response in advance, relying on a random generator to determine the action, delegating the decision-making responsibility to the user, or programming the vehicle to act in accordance with legal interests, depending on the circumstances[1881].

In such scenarios, it has already been established that programming based on conflicting legal interests fails to provide a legal solution in these situations. The use of random generators has also been deemed unacceptable. Furthermore, detecting dangers but refraining from taking preventive measures and leaving it to chance creates a void in responsibility[1882]. While delegating the decision-making responsibility to the individual

[1879] HILGENDORF, Automated Driving and the Law, 2017, p. 191.

[1880] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 156.

[1881] BECK, Das Dilemma-Problem, 2017, p. 136.

[1882] *Ibid*, p. 134.

in the driver's seat might appear to be a viable option, in practice, there is often insufficient time or chance for immediate actions of this nature[1883]. Additionally, there may be cases where no user is available to whom such a responsibility could be transferred to. In that case, in line with the prevailing opinion in German law, it may be argued that non-intervention (simply allowing events to take their course to avoid incurring liability) could be considered a valid option when faced with such dilemmas[1884]. However, avoiding programming altogether, such as by failing to install collision avoidance systems or accident algorithms, or the driver's disabling them, could itself constitute a basis for liability[1885]. This is because such systems are designed to minimise accident risks and mitigate harm, and are part of the duty of care[1886].

For instance, in a case where a self-driving vehicle is travelling through a narrow tunnel and calculates that continuing straight will certainly result in the death of one individual while swerving left poses a minor probability of killing another, what should be the programmer's course of action? It can be argued that, in such a situation, prioritising the option with the minor probability of causing harm is more appropriate both morally and under Section 34 of the StGB[1887]. This is because, in that case, the minor probability of death actually corresponds to a probability of injury, and one value substantially outweighs the other.

As can be observed from this instance, most of the dilemma examples presented in literature either overlook risk assessment (in terms of probability) or proceed based on the premise that one of the two outcomes will occur with certainty. Indeed, nearly all examples in dilemma scenarios focus on cases such as a sinking ship or a hijacked plane that must be shot down, where the outcome is portrayed as unavoidable and the decision directly determines the result. However, this perspective overlooks a critical point: these scenarios are thought experiments, and in real-life situations, such absolute certainty is seldom achievable.

In the event of a potential accident, a self-driving vehicle may decide to take action based on an assessment of the relative risk or harm posed by each option. Nevertheless, in practice, this may not yield the desired result. Even today's most sophisticated vehicles may fail to detect or accurately

[1883] SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 10.

[1884] GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 574.

[1885] HILGENDORF, Recht und autonome Maschinen, 2015, p. 22.

[1886] HILGENDORF, Dilemma-Probleme, 2018, p. 692.

[1887] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 156.

identify minor objects. For instance, a sudden braking manoeuvre might result in sliding, depending on the moisture of the road surface, which makes it challenging to accurately predict the outcome. However, it remains plausible that, in the future, more sophisticated self-driving vehicles will be capable of calculating such variables[1888].

The optimal course of action in programming self-driving vehicles is to establish a system which continuously monitors the environment to identify potential risks 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[1889].

In real-life scenarios, such as the frequently referred dilemma involving children suddenly running onto a road, it is highly unlikely that an isolated scenario devoid of all external factors and probabilities will occur. Instead, at the time that the children jump onto the road, a self-driving vehicle is far more likely to calculate complex probabilities. For instance, if a self-driving vehicle calculates that an accident is unavoidable and estimates a 40% likelihood of one person's death compared to a 98% likelihood for another, is it still possible to argue that both outcomes are equal? Or should it instead prioritise the option that would cause the least harm? What if the calculation were 98% versus 5%[1890]?

To illustrate further, at that moment, it might assess that continuing straight presents a 60% chance of the first child, who is 1.30 metres tall, being fatally struck, and a 95% chance of severe injury. If the vehicle slightly swerves to the right, the fatality risk for the first child drops to 30%, while the likelihood of hitting a curb and causing minor head injuries to self-driving vehicle's passengers rises to 35%, with a 5% chance of those injuries being fatal. Fully swerving right might raise the possibility of elderly pedestrians on the pavement failing to react to the manoeuvre and stepping into the vehicle's path to 25%, with a 10% chance of the car overturning, and an 80% likelihood of material damage. Conversely, swerving to the left could result in a 90% chance of injury and a 65% chance of fatality for the second child. At the same time, there is a 25% chance of colliding with an individual crossing on a bicycle, with a 5% probability of that collision being fatal. Moreover, even if the vehicle calculates that it can avoid killing one

[1888] LIN, Why Ethics Matters, 2016, p. 71.

[1889] HILGENDORF, Recht und autonome Maschinen, 2015, p. 23; HILGENDORF, Dilemma-Probleme, 2018, p. 692.

[1890] See: HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 161.

person by injuring another, the death of that person may still be inevitable. Such scenarios can be extended, highlighting the immense complexity and uncertainty involved in real-world moral dilemmas for self-driving vehicles. Therefore, it can be argued that choosing to swerve left or right does not, in real life, simply result in a choice between the death of one person and that of two; rather, it gives rise to far more complex outcomes.

In my view, the debates in literature remain overly reliant on classical moral dilemma thought experiments, often ignoring the probabilistic nature of real-life scenarios. In such situations, conduct that minimises risks should be prioritised. Furthermore, life vs. life dilemmas will be rare; instead, conflicts will typically involve legal interests of varying degrees[1891]. Additionally, such dilemmas are unlikely to arise suddenly and entirely unexpectedly. Self-driving vehicles can be programmed to anticipate the potential materialisation of a dilemma and act pre-emptively to prevent it[1892]. Indeed, limiting liability evaluation to the final moment of choosing between option A or B is, in my opinion, an inadequate approach. For example, 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[1893].

During lawful driving, situations such as the injury of a child who suddenly runs onto a road are typically assessed within the scope of permissible risk. However, when the same example involves two children instead of one, and completely avoiding a collision is impossible, the situation suddenly changes. In this context, an event that would ordinarily fall within the scope of permissible risk during lawful driving is reframed as intentional killing simply because, in the milliseconds available, the only possible action is to strike one child instead of two[1894]. This, in my opinion, is a flawed argument[1895].

[1891] BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 452 Rn. 48.

[1892] *Ibid*, p. 453 Rn. 50.

[1893] BECK, Das Dilemma-Problem, 2017, p. 133.

[1894] This real-life incident involves the sudden emergence of several animals and humans onto the roadway. At that moment, contrary to the claims of much of the literature, the vehicle does not encounter a genuine moral dilemma (although not a perfect example, it illustrates my point). Rather, it engages systems intended to avert an imminent collision. https://www.instagram.com/reel/DKo7V7uyQ9T. (accessed on 01.08.2025).

[1895] For a discussion on evaluating such situations within the framework of permissible risk, see: HILGENDORF, Recht und autonome Maschinen, 2015, p. 21.

For this reason, contrary to the majority of opinions in literature, I argue that the occurrence of isolated, pure dilemmas where intentional offences are at issue will be exceedingly rare. Instead, the focus should shift to examining most real-life situations through the perspective of negligence in conjunction with the duty to develop collision avoidance systems to the highest possible standard. In this context, the assessment of liability for collision avoidance systems designed to minimise risk should, without question, be conducted in parallel with the principles outlined under the concept of permissible risk.

The examination of such dilemmas through the perspective of permissible risk, particularly in relation to collision avoidance systems, has also been proposed in legal literature. *Hilgendorf* asserts that the determination of a manufacturer's liability in such dilemma scenarios ultimately hinges on whether a breach of the duty of care has occurred. He contends that this issue should be addressed within the framework of permissible risk[1896]. In scenarios where all individuals face equal danger from the outset, the vehicle should be programmed to minimise the number of innocent sufferers. However, the killing of innocent third parties remains unlawful, and the question of manufacturer liability remains unresolved. Nevertheless, if the manufacturer has taken all technically feasible and reasonable measures to prevent such emergency situations; the principle of permissible risk applies. In such cases, no negligence can be attributed, even if the vehicle causes harm or death to an innocent individual in a specific instance[1897]. However, in the context of sacrificing a life, the considerations emphasising the supreme value of life within the framework of necessity should not be overlooked[1898].

Similarly, *Schuster* argues that, since the emergency algorithms aim to minimise overall danger and reduce the likelihood of anyone becoming a victim, they benefit everyone and therefore may not create a legally disapproved risk, potentially excluding developers from liability[1899]. Indeed, from an *ex ante* perspective, causing harm to the fewest possible individuals

[1896] HILGENDORF, Autonome Systeme, 2018, p. 109.

[1897] HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 158; HILGENDORF, Dilemma-Probleme, 2018, p. 699; HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 169, 172 f.; HILGENDORF, Moderne Technik, 2015, p. 107, 110 f.

[1898] HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 164.

[1899] SCHUSTER, Das Dilemma-Problem, 2017, p. 114.

and minimising the number of accidents and damages represents the most reasonable scenario for all potential victims[1900].

Conversely, it has also been argued in literature that dilemmas cannot be resolved within the framework of permissible risk[1901]. Accordingly, scenarios such as the killing of an unrelated third party would surpass the limits of what is considered permissible and socially acceptable[1902]. Although the general systems in self-driving vehicles may be evaluated within the scope of permissible risk, dilemma scenarios where a conscious decision is made to sacrifice one individual fall outside this framework[1903].

Another critique comes from *Engländer*, who criticises *Hilgendorf*'s view for addressing dilemmas through the perspective of permissible risk. *Engländer* argues that permissible risk applies only to situations that are unavoidable despite the exercise of all due care. In contrast, in dilemmas, the violation of the legal interests of the specifically affected road users could be avoidable and preventable through alternative programming. Therefore, he contends that the concept of permissible risk is not applicable in such cases[1904]. However, it can be argued that *Engländer*'s critique is rooted in his interpretation of *Hilgendorf*'s arguments as being strictly tied to dilemmas, whereas *Hilgendorf* does not actually focus solely on dilemmas; but also addresses collision avoidance systems and risk minimisation.

Finally, it should be noted that classical dilemmas, where a definitive choice must be made between the lives of A and B, are possible; but will occur only in extremely rare circumstances. For all other situations, the explanations provided above under negligence and permissible risk remain applicable. Dilemma-like issues are instead more likely to arise in situations where AI-driven autonomous systems are used as decision-makers and must choose between multiple individuals (e.g. profiles). While the competing legal interests in such cases may not always involve life and death, they could instead pertain to equal or differing legal interests, such as property rights or other material claims.

[1900] According to Schuster, the matter should be resolved through the factual element of the crime and objective imputation. SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 11.

[1901] SEUFERT, Wer fährt, 2022, p. 329.

[1902] FELDLE, Notstandsalgorithmen, 2018, p. 89.

[1903] *Ibid*, p. 250

[1904] ENGLÄNDER, Das selbstfahrende, 2016, p. 375 ff., p. 388.
For Hilgendorf's response and counterarguments, see: HILGENDORF, Autonomes Fahren im Dilemma, 2017, p. 168 ff.

As autonomous systems become more widespread, dilemmas will increasingly arise in areas such as organ donation procedures[1905]. It is argued that employing chance (e.g., through a random generator) to make a decision is conceivable when choosing between two equally valuable legal interests, both of which cannot be saved -such as in cases where only one life-saving organ (*e.g.*, a heart) is available for two patients with identical tissue compatibility and waiting times on a transplant list. Unlike traffic-related dilemmas, there is nothing unlawful in deciding to allocate the heart to one patient over the other; however, failing to make any decision would result in the loss of a life and the waste of a transplantable heart[1906]. Furthermore, in terms of the applicability of existing legal constructs, there is no "right to an organ"; only a right to equal access to organ transplantation therapy[1907].

In conclusion, it should be noted that, it is of particular importance that critical decisions are made by humans rather than AI-driven systems. This is mainly to ensure accountability and moral responsibility; maintain transparency and trust; mitigate bias and error; incorporate empathy and contextual understanding, and enable adaptability in unique situations. However, even if a human ultimately makes a decision based on a report generated by an AI-driven system, the outcome is unlikely to differ significantly, as practical processes tend to follow a more pragmatic course. Moreover, due to the opacity of the machine's reasoning, it may not be possible to determine why it reached a particular (potentially biased) conclusion. Therefore, future academic research may prove more constructive if it directs greater attention to these contexts rather than on self-driving vehicles; where concepts such as necessity as a justification and exculpation, as well as supra-legal necessity and conflict of obligations, could be applied.

[1905] HILGENDORF, Dilemma-Probleme, 2018, p. 682.

[1906] JOERDEN, Zum Einsatz, 2017, p. 88 f.

[1907] SCHUSTER, Das Dilemma-Problem, 2017, p. 109.