Summary

In criminal offences involving AI-driven autonomous systems, significant challenges arise in determining the liability of the person behind the machine. Rather than focusing on a specific AI application, this study seeks to establish a general framework aimed at delineating and concretising the scope and boundaries of liability, particularly in cases of negligence. In this context, certain observations and insights may be proposed:

When examining liability, emphasis should be placed on 'autonomy' rather than 'artificial intelligence'; as the key concern lies in the technical autonomy of these systems, the diminished human control, and the unpredictability of their outcomes. In addition to such *ex ante* challenges, *ex post* difficulties arise in determining the causal nexus for liability. They stem from the opacity of such systems, which may result from algorithmic confidentiality, the general public's limited technical expertise, and the complexity of managing extensive datasets and parameters.

To address potential liability gaps, the legal literature has extensively debated granting robots personhood and assigning their own liabilities. This perspective, rooted in an anthropomorphic view, overlooks the fact that AI systems inherently lack genuine moral reasoning, a will and the capacity to understand their conducts. Unlike corporate liability, this approach encounters numerous technical challenges, such as the inability of AI systems to perform acts that are relevant under criminal law. Consequently, this form of liability cannot be explained through analogies but can only be addressed through serious legal fictions based on pragmatic necessities. Such an approach is unlikely to be feasible in the near future, particularly under current legal frameworks.

Unlike criminal law, civil law mechanisms such as strict liability can somewhat simplify the determination of liability. While some functions of criminal law in ensuring justice and social order can partially be addressed through civil law mechanisms, the two legal branches serve fundamentally different purposes, and civil law liability models are not adaptable to criminal law. Consequently, rather than a criminal liability gap, a retribution gap will emerge. In the future, as such systems become more widespread, it will be necessary to assess whether living with such a gap would align with the expectations placed upon the legal order by society.

https://doi.org/10.5771/9783748965183-443 https://www.inlibra.com/de/agb - Open Access -

Criminal product liability may be applicable to manufacturers of AI-driven autonomous systems. However, three key challenges arise: defining these systems as products, identifying what constitutes a defect in their context, and addressing the burden of proof difficulties stemming from their opacity.

AI-driven autonomous systems do not exhibit significant differences regarding intentional crimes; liability is determined as long as the causal nexus can be established. On the other hand, contrary to the part of the literature, the indirect perpetration model cannot be applied to crimes where such systems are utilised; because they lack will, their conduct does not qualify as an act under criminal law, and they cannot be regarded as "another" in the human sense.

In the context of negligent liability, the duty of care derives from a multifaceted framework encompassing statutory legal provisions, codes of conduct, behavioural standards, professional guidelines, administrative and operational instructions, usage protocols, and unwritten norms. However, compliance with these standards serves only as an indicator; general principles, such as the duty to refrain from harm, remain applicable in all cases. Ultimately the determination of negligence is made by the court, considering all the specific circumstances of the case.

In cases of negligence, an individual's / organisation's specific knowledge and skills are taken into account. *E.g.*, if a company has developed a method to reasonably reduce risks, it must be implemented, even if it has not yet become an industry standard. Additionally, individuals or organisations that engage in risky activities despite lacking the capacity or expertise to manage the associated dangers are held liable for harmful outcomes under negligent undertaking.

For liability in negligence, the harmful outcome must be at least generally foreseeable and avoidable. However, the risks posed by AI-driven autonomous systems are themselves recognisable. Therefore, the liability of individuals who delegate tasks to such systems, instead of performing them through conventional methods, should be examined. This does not imply that the individual will be liable in all cases. Rather, it necessitates a detailed examination by recognising the delegation of the task as an act within the meaning of criminal law. Consequently, the widely recognised view in the literature, which considers such individuals merely passive and therefore not liable, is open to criticism. Nonetheless, a distinction must be made between typical and atypical risks in such cases. Moreover, the duty

https://doi.org/10.5771/9783748965183-443 https://www.inlibra.com/de/agb - Open Access -

of care is further shaped by lessons derived from past incidents and the new possibilities enabled by technology.

The complete elimination of risks associated with AI-driven autonomous systems is not feasible, and the permissible risk doctrine guides the assessment of the duty of care. In the absence of established experience and standards, the state of science or technology may need to be applied to mitigate risks to a permissible level. This approach aligns with the dynamic nature of the field. Identifying which activities qualify as permissible risks is challenging to determine *ex ante*. While standards may alleviate some of the pressure on actors, they cannot provide complete relief, as they function merely as indicators. Pre-compliance through formal *box-ticking* does not grant actors a *carte blanche*. Ultimately, the focus remains on whether the necessary measures to reduce risks were appropriately implemented.

The risks associated with AI-driven autonomous systems may be deemed permissible if all necessary measures are undertaken to reduce such risks to an acceptable level, and if these risks align with the degree of societal tolerance. In this context, societal gains and potential risks must be carefully evaluated. In this assessment, if, as suggested in the literature, general considerations unrelated to specific tasks (such as economic and environmental contributions) are taken into account, the overall negative impacts must also be considered. Furthermore, it is not possible to make a general approach for all AI applications. In this regard, a calibration model should be implemented to mitigate risks to a permissible level, taking into account the risks (severity and likelihood of harm) posed by the activity in question, as well as the functions it serves within society. Thus, risky activities that benefit only certain segments of the society, and activities which are indispensable for the society should not be evaluated equally, and a measured duty of care appropriate to the nature of the activity can be ensured.

Whether delegating a task to AI-driven autonomous systems enhances the risks compared to performing the task using conventional methods should be examined. However, risks and hazards are not merely quantitative variables that increase or decrease; rather, they involve a form of substitution. In specific cases, certain hazards may intensify while others diminish. In any case, an evaluation can be conducted based on the risk enhancement theory (*Risikoerhöhungstheorie*). This approach ultimately serves to prevent individuals who transfer the risks and responsibilities of an activity to autonomous systems, thereby placing themselves in a "passive" position, from exploiting the concept of permissible risk.

https://doi.org/10.5771/9783748965183-443 https://www.inlibra.com/de/agb - Open Access -

If performing a task through AI-driven autonomous systems significantly reduces risks, introduces no novel or unacceptable risks, and is socially accepted, the failure to use such systems in the future could constitute a breach of the duty of care.

Although the "EU AI Act" does not directly address criminal liability, it imposes certain requirements and obligations on relevant parties based on the level of risk associated with AI. These provisions can serve as a reference for defining the duty of care under national law.

The development, deployment, and use of AI-driven systems often involve multiple parties, and the issue may arise either from the actions of one individual or from a combination of them. In this regard, the matter does not significantly differ from classical criminal law (*e.g.* product liability) cases and the principle of reliance applies with its limitations.

Extending the principle of reliance to AI-driven autonomous systems presents certain challenges. First, individuals are typically subject to monitoring obligations to ensure that these systems function correctly. On the other hand, machines must be designed to account for foreseeable and often typical human errors. Moreover, the principle of reliance is a concept developed for humans, grounded in their biological capacities. In contrast, machines, through their sensors and data processing capabilities, can perform continuous monitoring. Therefore, the principle of reliance does not need to be applied to machines in its original form.

Contrary to the prevailing view, self-driving vehicles are unlikely to encounter pure typical dilemma scenarios. In this regard, the use of *state of the art* collision avoidance systems should be assessed under the concept of permissible risk within the context of the duty of care. In rare cases where such a pure dilemma arises, the necessity as exculpation or justification, as well as conflict of obligations, fail to provide a satisfactory resolution. The application of supra-legal necessity, on the other hand, has been subject to various criticisms in the literature. Nonetheless, the principle that life holds the highest value must remain inviolable.

As has been proposed in the literature for *de lege ferenda*, stipulating the placement of dangerous products on the market without adequate safety measures as an abstract endangerment offence, with the occurrence of harm serving as an objective condition of punishability, offers a reasonable framework for deterrence by addressing many of the challenges in determining criminal liability. However, this approach also encounters challenges and raises certain concerns due to the specific characteristics of AI.

https://doi.org/10.5771/9783748965183-443 https://www.inlibra.com/de/agb - Open Access -