Chapter 3: Doctrinal Approaches to Liability Models in the Literature

A. Bridging Contested Liability Gaps in Criminal Law

Criminal liability in cases involving autonomous systems, particularly those driven by AI, poses significant challenges due to their inherent autonomy and opaque nature. Therefore, as discussed in detail above[284], attribution of liability is complicated by the level of autonomy these systems possess compared to traditional systems. This often leads to a debate regarding a liability gap in criminal law doctrine, which existing legal frameworks struggle to address adequately. AI-driven autonomous systems may cause violations not only under criminal law, but also within administrative and civil law. While these systems pose challenging issues in civil law as well, certain established approaches provide clearer pathways for determining liability, making it comparatively easier to address. However, criminal liability fundamentally rests on an individual's culpable violation of a penal norm that protects legal rights or interests. This raises complex questions: are current criminal law principles sufficient for addressing present and future challenges? Can AI-driven autonomous systems be granted legal personhood for practical reasons and held liable? What level of due care and foreseeability should be legally expected for persons behind the machine? Could and should crimes involving these systems go unpunished if no blameworthy party is found?[285] The complex nature of AI complicates the assessment of causality and the attribution of liability, creating what some scholars describe as a contested "gap". Several solutions have been proposed in scholarly literature to bridge this gap and address the unconventional cases involving these systems to adapt existing liability models.

Particularly in the field of criminal law, tracing the source of influential ideas that have shaped discussions and even impacted views within Conti-

284 See: Chapter 1, Section E: "Distinctive Challenges of Crimes Involving AI-Driven Autonomous Systems".

285 In the future, as fully autonomous vehicles become widespread and drivers are relieved of their duty of supervision, the occurrence of situations where no one can be held criminally liable is expected to increase. WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1122.

nental Europe reveals that they are primarily based on the works of *Gabriel Hallevy*, a legal scholar, relying predominantly on the Anglo-American approach in criminal law[286]. *Hallevy* proposes mainly three liability models: *perpetration-by-another liability model*, *natural-probable-consequence liability model* and *direct liability model*[287]. These models have been extensively discussed by various criminal law scholars[288] and have been supported by some. Moreover, even if these models have not been evaluated directly by referring to *Hallevy*'s early publications, various studies have advocated for the application of one of these models, namely the "perpetration by another" model, in cases involving the utilisation of robots as instruments. This indicates that his works have been highly influential in legal literature.

The liability models discussed in literature extend well beyond these examples. Numerous alternative models have been put forward by drawing parallels between the characteristics of robots and familiar human concepts. For instance, as early as 1981, various liability models for artificial agents were proposed, including analogies to dangerous animals, slavery, product liability, diminished capacity, children, agency and personhood[289]. However, as examining all these models falls beyond the scope of this study, only the most prominent ones will be discussed, and their potential adaptation to address possible liability gaps in criminal law will be assessed. Subsequently, in *Chapter 4*, solutions will be sought within the framework of traditional criminal law doctrine.

286 STRASCHNOV, The Judicial System in Israel, 1999, p. 527 ff.

287 HALLEVY, The Criminal Liability, 2010, p. 174; HALLEVY, When Robots Kill, 2013, p. 64 ff.

288 FREITAS/ANDRADE/NOVAIS, Criminal Liability of Autonomous Agents, 2014, p. 149 f.; KING, et al., Artificial Intelligence Crime, 2020, p. 108; PAGALLO, From Automation to Autonomous Systems, 2017, p. 19; MAHMUD, Application and Criminalization, 2023, p 9 f.; VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 664; DOBRINOIU, The Influence, 2019, p. 144.

289 LEHMAN-WILZIG, Frankenstein Unbound, 1981, pp. 447-453.
For a similar study conducted in 2012, see, *inter alia*: ASARO, A Body to Kick, 2012, pp. 170-180.

B. Autonomous System's Own Liability

1. Fundamentals

Among the proposed liability models, perhaps the most debated, mainly influenced by the long-standing *sci-fi* culture, is the liability of a robot (AI-driven autonomous system) itself. The advancement of AI-driven robots has led to their deeper integration into daily life, shifting the perception of robots from mere possessions to more human-like entities and moving from a traditionally anthropocentric perspective to an anthropomorphised approach[290]. Although this topic may appear novel, it was in fact addressed nearly half a century ago[291], and even before. The underlying rationale is that a gap in criminal liability[292], or even the mere perception of such a gap in society, results in undesirable consequences and hinders criminal law from fulfilling its purpose. Therefore, it is argued that the criminal liability of robots must be thoroughly considered[293]. Particularly as autonomy increases, it will become more reasonable to consider the notion of a robot's own responsibility in the future[294].

To discuss the concept of a robot's own liability, three main legal issues arise under *de lege lata*. First, from a legal standpoint, the robot must be capable of performing an act to provide a basis for examining criminal liability. Secondly, they must possess culpability; a guilty mind[295]. Thirdly, they must be suitable subjects for a conviction or the imposition of a criminal penalty[296].

The introduction of direct liability for AI-driven autonomous systems hinges on their recognition as independent subjects of legal relations[297].

290 DEHNERT/GUNKEL, Beyond Ownership, 2023, p. 6 ff.

291 LEHMAN-WILZIG, Frankenstein Unbound, 1981, p. 443.

292 The issue of whether there is a criminal liability gap is contested. Setting aside future possibilities, it is a widely held view that current criminal law is mostly adequate for addressing and categorizing cases involving AI without significant responsibility gaps. See *inter alia*: SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 513.

293 QUARCK, Zur Strafbarkeit, 2020, p. 66.

294 LIN/ABNEY/BEKEY, Robot Ethics, 2011, p. 946.

295 SWART, Constructing Electronic Liability, 2023, p. 595.

296 QUARCK, Zur Strafbarkeit, 2020, p. 66.
For instance, due to these criteria, the author does not accept the "intelligent agents" as persons, but mere tools or machines from a legal aspect. See: SEHER, Intelligent agents, 2016, p. 60.

297 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 667; ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 383.

The criminal liability of legal persons other than natural persons varies across different legal systems. For instance, corporate criminal liability has been recognised in jurisdictions such as the USA, the UK, Austria, France, the Netherlands, Portugal, Spain, Switzerland, and in international criminal law. In Germany, however, corporate criminal liability is not granted, based on the underlying premise that legal persons do not possess culpability; instead, they can only be subjected to administrative fines[298]. Nonetheless, a legal system is free to hold non-human actors liable. It has been argued that once this conceptual hurdle is overcome, attributing liability to robots would not be particularly difficult[299].

Under Turkish law, criminal liability of legal persons is not recognised; however, according to Art. 20(2) of the Turkish Penal Code (TPC), security measures can be imposed on them in connection with criminal offences. Therefore, it can be argued that the capacity of legal persons to perform acts is, to some extent, acknowledged by the legislator, as security measures are sanctions prescribed in response to criminal acts[300]. However, a counterview argues that the imposition of security measures on legal persons does not necessarily imply that they should be considered as the entity performing the criminal act; because legal persons do not possess the capacity to act, and consequently, they inherently lack the capacity for culpability[301]. However, it is explicitly stated in Article 49 of the Turkish Civil Code[302] that legal persons also possess the capacity to act which they can perform through their organs. The relationship between the organ and the legal person is not one of representation[303].

Particularly in legal systems rooted in common law, the established practice of assigning criminal liability to corporations supports the idea of extending such liability to robots without further rationale. However, even some perspectives that do not oppose the concept of direct criminal liability for robots, challenge this default assumption and advocate for

298 KINDHÄUSER/ZIMMERMANN, § 7 Handeln für einen anderen - Strafrecht AT, 2024, p. 71 Rn. 1 fn. 1. There are differing views on whether legal persons possess the capacity to act through their organs. For instance, one view denies such capacity: CORNELIUS, Künstliche Intelligenz, 2020, p. 61.

299 HILGENDORF, Können Roboter schuldhaft handeln?, 2012, p. 127.

300 KATOĞLU, Ceza Hukukunda, 2012, p. 667.

301 ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 213 f.

302 Turkish Civil Code No. 4271, dated 22.11.2001 (Official Gazette No: 24607, 08.12.2001)

303 KATOĞLU, Ceza Hukukunda, 2012, p. 668 ff.

the evaluation of the necessity of substantial justification as a preliminary step[304].

The crimes committed by legal entities, such as companies, generally have a financial aspect or involve issues such as environmental pollution[305]. Therefore, with regard to AI-driven systems, there are significant conceptual challenges when considering offences such as homicide or bodily harm. Indeed, attributing liability to the owner or supervisor is more reasonable within the framework of existing legal notions concerning today's autonomous entities[306].

Criminal law, unlike civil law, requires that an offence be committed by a moral agent. While harm can occur without moral agency, there can be no guilt without a guilty mind[307]. One of the main arguments supporting the idea that corporations cannot commit crimes (*societas delinquere non potest*) is that they are incapable of guilt. However, this concept is not, in fact, foreign to the civil law tradition and was not always consistently applied within the context of Continental European law. Initially, corporate criminal liability was recognised in both common law and civil law traditions. Nonetheless, with the advent of *Enlightenment* and the emphasis on the principle of individual guilt, corporate criminal liability was eventually abolished in German law[308].

Particularly within Western philosophical traditions, humans are considered moral agents because they possess the ability to freely choose their actions and abstain from others. Although some perspectives tend to anthropomorphise computers and treat them as if they were moral agents, the prevailing consensus among most philosophers is that current computer technologies should not be viewed as moral agents[309]. Indeed, given the current state of technology, it can be stated with confidence that attributing culpability to AI-driven autonomous systems is not feasible; unless a fundamentally new concept of guilt, differing significantly from

304 HU, Robot Criminals, 2019, p. 492.
305 VAN DEN HOVEN VAN GENDEREN, Do We Need Legal Personhood, 2018, p. 34.
306 REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 74.
307 ASARO, A Body to Kick, 2012, p. 181.
308 DUBBER, The Comparative History, 2013, p. 204 ff.
309 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).

traditional notions of free-will[310], freedom and its execution is developed[311]. Nonetheless, some scholars argue that, in the future, AI systems may develop human-like characteristics or achieve such complexity that they fulfil normative expectations and could potentially be regarded as entities capable of bearing responsibility[312]. It is noted that, stemming from a robot's own strict liability, sanctions such as banning its use or correcting system flaws could be contemplated, incorporating principles from administrative law and related sanctions[313].

It has been discussed that, in medieval Europe, animals were sometimes personified as incarnations of dark forces and punished in ways similar to humans, such as hanging, crucifixion or burning; motivated by retribution which is the essence of penal sanctions. Additionally, injured parties could claim the animal as compensation[314]. However, contrary to popular belief, formal 'criminal' trials for animals with human-like sentences were likely rare and typically ended with the animal being killed as a precaution. Furthermore, using this as an argument for robotic responsibility is considered absurd[315].

Determining appropriate sanctions for non-real persons and their functionality involves complex topics related to the dogmatics of criminal law and sanctions which go beyond the scope of this study. Despite the assertion that atonement and preventive effects of punishment do not apply to legal persons and are only relevant to natural persons[316], sanctions of any kind can have a deterrent effect on both natural and legal persons. If justice is believed to be achievable only through *inter alia*, retribution, sanctions such as the destruction or reprogramming[317] of AI-driven autonomous systems in response to a serious malfunction might be seen not only as serving general and specific preventive purposes but also as a form of retribution. However, such sanctions would not serve the functions of a

310 The topics of guilt and free will extend well beyond the scope of this study. However, for a discussion on intelligent agents and related debates, see: GLESS/WEIGEND, Intelligente Agenten, 2014, p. 573 ff, 579.
311 JOERDEN, Strafrechtliche Perspektiven, 2013, p. 205.
312 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 260.
313 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 667.
314 BREDNICH Rolf Wilhelm, Enzyklopädie des Märchens, 2010, pp. 649-654; GLESS/WEIGEND, Intelligente Agenten, 2014, pp. 566-567.
315 FISCHER, Gefährliche Sachen, 2020, p. 128 fn. 1; SCHUSTER, Künstliche Intelligenz, 2020, p. 393.
316 ÖZGENÇ, Türk Ceza Hukuku, 2019, p. 213 f.
317 BESTER, The Demolished Man, 1978, p. 237 ff.

criminal penalty[318] as current software lacks the capacity for volition or the ability to comprehend sanctions[319].

2. The Legal Debate on Personhood for AI-Driven Autonomous Systems

a. Pro Arguments in Legal Literature for AI-Personhood

(1) The Origins

The recognition of personhood for AI-driven systems has been a topic of extensive debate for a considerable period, particularly in the context of potential legal issues that may arise[320]. One of the earliest contemporary suggestions related to the topic can be found in the 2012 report of *euRobotics*[321]. Following extensive discussions, the European Parliament's 2017 recommendation to the Commission[322] for the introduction of an "electronic person" has been important in reviving debates on legal personhood to address liability gaps[323]. According to this proposal, advanced autonomous robots would eventually be assigned electronic personhood, making them

318 Still, as long as there is a difference between killing a human being and formatting a hard drive, the idea of punishing machines will remain a misleading use of the term. See: ROXIN/GRECO, § 8. Handlung in Strafrecht AT, 2020, p. 370 Rn. 66 f.

319 SCHUSTER, Das Dilemma-Problem, 2017, p. 103.

320 SOLUM, Legal Personhood for AI, 1992, p. 1284 ff.

321 Exploration track: non-human agents and electronic personhood, Suggestion for a green paper on legal issues in robotics, Eds.: LEROUX C./LABRUTO, R., euRobotics The European Robotics Coordination Action, 2012, https://www.researchgate.net/publication/310167745_A_green_paper_on_legal_issues_in_robotics, pp. 58-64. (accessed on 01.08.2025).
See also: GÜNTHER, et al., Issues of Privacy and Electronic Personhood in Robotics, 2012, p. 819 f.
In fact, the debates date back much further. However, discussions focused on concrete actions are relatively recent. For instance, regarding a debate from 2007, see: TEUBNER Gunther, "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law", Max Weber Lecture Series MwP 2007/04, 17.01.2007, https://hdl.handle.net/1814/6960, p. 20. (accessed on 01.08.2025).

322 European Parliament, Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), Committee on Legal Affairs, A8-0005/2017, 27.01.2017 https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.pdf, p. 7, 18. (accessed on 01.08.2025).

323 SWART, Constructing Electronic Liability, 2023, p. 596; CORNELIUS, Künstliche Intelligenz, 2020, p. 53; KIZILIRMAK, Yapay Zekâlı, 2021, p. 12 f.
For the earlier debate, see: BECK, Über Sinn, 2013, *passim*.

liable for any damage they might cause, especially in relation to compensation claims. Although this proposal was predicated on the assumption that AI-driven autonomous systems would become more sophisticated over time; in the past several years, many experts have criticised the notion, arguing that such entities have not yet reached that level of advancement. As a result, debates on electronic personhood have been set aside for the time being, without the formulation of a legal framework.

(2) Anthropomorphising Robots

Owing to advancements in AI, digital systems have increasingly assumed tasks traditionally reliant on human intellectual capacity, including computation, decision-making and control. Consequently, these systems, unlike other inanimate objects, are often attributed with mental characteristics such as intention and preference[324]. As robots increasingly resemble humans, the notion of categorising them solely as "things" has begun to appear less appropriate[325]. The notion of granting machines human-like status strengthens as their daily interactions with people increase[326] and eventually, the distinction between "human" and "person" may become blurred[327]. It is therefore argued that we are on the edge of introducing a new legal category that bridges the line between personhood and objecthood, necessitating adjustments within legal frameworks to accommodate this development[328].

Anthropomorphic perspectives, also referred to as "android fallacy"[329], go further by arguing that morality should not be confined to human agents but should also include artificial agents. Despite lacking consciousness, these agents, with their ability to interact with the environment, act autonomously, adapt to new situations, and can develop a form of moral

324 NISSENBAUM, Accountability in a Computerized Society, 1996, p. 35.
325 MULLIGAN, Revenge Against Robots, 2018, p. 594.
326 HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 549.
327 SOLUM, Legal Personhood for AI, 1992, p. 1260.
328 CALO, Robotics and the Lessons, 2015, p. 549.
329 The term describes the erroneous attribution of human-like qualities to robots, leading to potential legal and ethical misjudgements due to the anthropomorphic perception of these machines as autonomous entities with moral agency. RICHARDS/SMART, How should the law, 2016, pp. 18-21.

responsibility through learning and feedback[330]. Even the criticism that it is too early for such discussions is rejected by some, as current advancements suggest that robots capable of moral decision-making will evolve rapidly[331]. It is even argued that future AI-driven systems, when equipped with specific technical attributes and granted legal personhood, could fulfil the *mens rea*[332]. Accordingly, it is suggested that when AI-driven systems possess the capability to meet the awareness requirements in criminal law -though not referring to human-like awareness- they could commit offences both intentionally and negligently[333]. This implies that general defences in criminal proceedings, such as loss of self-control, insanity, intoxication or factual and legal mistakes, could potentially be applied in favour of artificial agents[334]. Moreover, it would be possible for robots to be held liable not only as direct perpetrators but also as accomplices, joint perpetrators, inciters, or accessories[335].

The argument that AI-driven systems should be granted personhood has been advocated on the grounds that there are precedents for such a decision, with examples such as New Zealand courts recognising certain natural entities like rivers; and Argentina granting legal personhood to an orangutan named Sandra[336]. However, it should be noted that, aside from potential misunderstandings in these examples, there are significant differences within the concept of legal personhood. While in common law tradition, attributing personhood status to things such as machines can be more easily justified[337], this is less feasible in Continental Europe. Due to its intellectual history rooted in theological and philosophical backgrounds since the Enlightenment, such recognition is more difficult to achieve. Despite the technical autonomy that robots may exhibit, they remain machines, and are therefore classified as "things"[338].

The potential solution of recognising a different status, such as personhood for robots rather than that of mere "things", to fill the liability gap

330 FLORIDI, On the Morality of Artificial Agents, 2004, p. 375; SØVIK, How a Non-Conscious Robot, 2022, p. 797.
331 HU, Robot Criminals, 2019, pp. 492-493.
332 MÜSLÜM, Artificial Intelligence, 2023, p. 176.
333 HALLEVY, Liability for Crimes Involving AI, 2015, p. 124 ff.
334 PAGALLO, From Automation to Autonomous Systems, 2017, p. 19.
335 HALLEVY, Liability for Crimes Involving AI, 2015, p. 104 ff.
336 SWART, Constructing Electronic Liability, 2023, p. 597.
See also: TUNÇ, Can AI Determine, 2024, *passim*.
337 VLADECK, Machines Without Principals, 2014, p. 124.
338 HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 548 ff.

raises numerous legal issues that must first be clarified[339]. The foremost among these is the determination of the nature of the status to be conferred[340] and the specific rights that would be attached to it. In literature, a number of potential statuses have been put forth for consideration. One approach is to view robots as property, a purely legal and moral object. Alternatively, they may be regarded as messengers or representatives, with a specific legal status. Another option is to view robots as indirect rights-holders, a status as suggested similar to that of animals. They could also be regarded as having specific rights and duties, akin to the current status of legal persons. Finally, robots could be viewed as having comprehensive rights and duties, comparable to the status of natural persons. If -hypothetically- rights are to be granted to robots, it is imperative that these rights are tailored to their unique nature[341]. For instance, it would be erroneous to assume that robots possess expectations of privacy or dignity[342].

(3) Pragmatical Necessities

Granting personhood to AI-driven autonomous systems should be approached from a legally pragmatic and necessity standpoint, rather than from anthropomorphic perspectives that suggest robots meet certain human-like conditions. The determination of the criteria for the recognition of legal personhood is, indeed, a complex matter. While the will of individuals such as infants or those in a vegetative state may be open to debate, they are unquestionably legally considered natural persons. If social interaction were to be the criteria, many intelligent animals could also qualify as examples[343]. Therefore, the key factor for creating such a legal fiction may only lie in pragmatic necessities.

It has been argued that the law, which has already expanded the concept of personhood to include non-human entities such as corporations that lack physical existence[344], would not face significant difficulty in granting legal personhood to machines; since robots can directly interact with hu-

339 HILGENDORF, Recht und autonome Maschinen, 2015, p. 28.
340 See: Chapter 3, Section C(1)(b)(2): "Exploring Existing Frameworks: Slavery, Animal Ownership, Employees and Associates".
341 For the potential status, see: BECK, Über Sinn, 2013, p. 252, 255.
342 TURNER, Regulating AI, 2019, pp. 170-171.
343 TUNÇ, Legal Personhood for AI, 2022, p. 576.
344 SCHUPPLI, Can Legal Codes, 2014, p. 4.

mans in the physical world[345]. Nonetheless, while personhood is necessary for attributing criminal liability, it is not sufficient on its own[346].

The rationale behind having multiple categories of legal status lies in the practical necessity of addressing the varying levels of importance, rights and responsibilities that different entities have, both from legal and moral perspectives[347]. Moreover, law is inherently a discipline that operates on assumptions and fictions; for example, the entire legal system is constructed on the presumption of free will. In this context, constructing criminal liability on guilt -defined as the injustice resulting from the commission of a crime that disrupts the social order- rather than on free will, can be seen as a more reasonable approach. Redefining the concept of guilt in this functional manner enables the attribution of liability -and as a prerequisite, personhood- to intelligent agents[348]. In this context, it is emphasised that a pragmatic approach should be adopted in law. If a point is reached where AI-driven systems make autonomous decisions and perform tasks similarly to humans, the legal definitions of 'person', as well as concepts such as crimes of intent, negligence, and strict liability, could be radically redefined[349].

The concept of 'person' in law is not static, but dynamic depending on practical reasons[350]. Identifying the responsible person behind the machine is becoming an increasingly difficult task, potentially due to both intentional and unintentional complications. To mitigate this challenge, recognising personhood for AI-driven autonomous systems addresses a practical need, allowing for the acceptance of the machine's own (strict) liability[351]. Such recognition of personhood would eliminate ambiguities and enhance legal certainty[352]. Furthermore, this approach would be sensible not only from a criminal law perspective but also from a broader legal policy standpoint[353].

345 SWART, Constructing Electronic Liability, 2023, p. 602; BECK, Intelligent Agents and Criminal Law, 2016, pp. 141-142; ALTUNÇ, Yapay Zekâ, 2021, p. 364; VAN DEN HOVEN VAN GENDEREN, Do We Need Legal Personhood, 2018, p. 35 f.
346 KÖKEN, Yapay Zeka, 2021, p. 263, 271-272.
347 BECK, Über Sinn, 2013, p. 245.
348 QUARCK, Zur Strafbarkeit, 2020, p. 68.
349 PAGALLO, From Automation to Autonomous Systems, 2017, p. 19.
350 HILGENDORF, Recht und autonome Maschinen, 2015, p. 16.
351 SWART, Constructing Electronic Liability, 2023, p. 594.
352 *Ibid*, p. 599.
353 QUARCK, Zur Strafbarkeit, 2020, p. 68.

(4) Defining the Nature and Scope of Legal Personhood for Robots

Should the concept of conferring electronic personhood upon robots be accepted[354], the question of which entities should be recognised will inevitably arise. Some scholars argue in literature that personhood should only be conferred on AI-driven systems if they achieve a level of self-awareness[355] or complete autonomy[356]. Others advocate recognising personhood for highly sophisticated embodied systems as well as software agents[357]. It is also argued that only highly advanced AI systems with a physical presence in the external world, equipped with actuators, could be considered for personhood and criminal liability, whereas software agents, lacking such physical embodiment, are excluded from this consideration[358].

A significant debate surrounds the question of whether robots must be moral agents[359] for their personhood and direct liability to be acknowledged. This issue, however, encompasses a range of metaphysical and philosophical aspects. It has been suggested that non-human entities can function as moral agents. Remarkably, in the U.S., judicial authorities have imposed liability on various legal persons for offences, even when individual human representatives were not personally culpable[360].

A view based on ethical behaviourism holds that the observable behaviour of robots should guide our ethical treatment of them. Moral consideration should be extended based on their behaviour and capacities rather than their intrinsic characteristics: if they appear sentient, capable of suffering, or show other morally relevant traits, they should be treated accordingly. Consequently, if they resemble humans, they should be treated as such[361].

Another view suggests that criminal liability for robots could apply only to 'smart robots' which are moral agents. Accordingly, smart robots are equipped with algorithms capable of making significant morally relevant

354 BECK, Intelligent Agents and Criminal Law, 2016, p. 141 ff.
355 AKSOY, Yapay Zekalı, 2021, p. 24.
356 VLADECK, Machines Without Principals, 2014, p. 124.
357 SWART, Constructing Electronic Liability, 2023, p. 596.
358 KÖKEN, Yapay Zeka, 2021, p. 272.
359 According to the weak notion of agency, AI systems can be considered agents if they include autonomy, social ability, reactivity, and pro-activeness. WOOLDRIDGE/JENNINGS, Intelligent Agents, 1995, p. 116.
360 HU, Robot Criminals, 2019, p. 517.
361 DANAHER, Welcoming Robots, 2020, p. 2025 ff. For the assessment of the view: MAMAK, Robotics, 2023, p. 34

decisions, can communicate these moral decisions to humans, and are allowed to act in their environment without immediate human supervision[362].

Despite contrary views[363] it has been widely argued that machines cannot fulfil *mens rea* and therefore, from *a de lege ferenda* perspective, only their criminal strict liability can be recognised[364]. In response, it is argued that certain advanced robots, such as "smart robots", do not require the pursuit of intention or guilt in the traditional sense of morally wrongful conduct. Based on their programming, they can assess that their conduct is wrong and recognise that a moral principle applies to a given situation, allowing them to understand that their conduct is wrong[365]. Consequently, if advanced robots of the future are granted personhood and recognised as moral agents, their guilty mind could be assessed, resulting in potential punishment. Their criminal liability would be no different from that of humans[366].

Artificial intelligence-driven embodied systems which exhibit a certain level of autonomous behaviour and are specifically designed for social interaction with humans and lifelike responses to mistreatment -referred to as "social robots"- are also argued to be moral agents and should be protected under specific laws[367]. It is also argued that even if the moral status of robots is not recognised, their significance demands protection through separate criminal norms[368].

In a similar manner to the assignment of criminal liability to humans only upon attaining a certain level of life experience and volitional development, such as by the age of 15, it can be argued that only robots that have reached a sufficient level of sophistication can be considered moral agents. Responsibility is therefore seen as a matter of degree rather than an absolute. Furthermore, the application of criminal sanctions to robots can be viewed as a form of feedback, guiding them to choose correctly[369].

362 HU, Robot Criminals, 2019, p. 490, 502.
363 MÜSLÜM, Artificial Intelligence, 2023, p. 176.
364 SWART, Constructing Electronic Liability, 2023, p. 598 ff.
365 HU, Robot Criminals, 2019, p. 522 f.
366 To speak of a guilty mind, the entity in question must first possess the capacity to act otherwise. See: SIMMLER/MARKWALDER, Guilty Robots?, 2019, p. 10, 27.
367 DARLING, Extending legal protection, 2016, p. 228.
368 MAMAK, Robotics, 2023, p. 35.
369 SØVIK, How a Non-Conscious Robot, 2022, p. 797.

A potential legal framework for personhood (electronic personhood) could involve the establishment of a liability fund, contributed by all stakeholders (programmers, manufacturers, sellers and users), proportional to the machine's risk, application and autonomy, which could grow over time through the robot's activities. The fund would cover damages clearly caused by the machine in cases where no human fault can be proven. Additionally, electronic persons should be registered in a system similar to a commercial register, with a unique number to allow those interacting with the machine to assess associated risks[370].

(5) The Impact of Robotic Liability on the Responsibility of the Person Behind the Machine

The potential impact of recognising robots as legal persons with liability on the responsibility of individuals associated with them (the person behind the machine) is significant. In certain cases, particularly with regard to civil liability, it could limit or even preclude the liability of these individuals[371]. However, this reasoning does not align with the core principles of criminal liability, which would still require holding those individuals accountable.

Assigning criminal liability directly to robots would not preclude the criminal liability of the persons behind the machine, assuming that their culpability can be proven[372]. Due to the principles of individual criminal responsibility and guilt, anyone who is at fault would be held liable under criminal law, provided that the other necessary conditions are met[373]. Especially, a person whose negligent behaviour contributes to a system's malfunction would continue to bear criminal liability[374]. Furthermore, in cases involving advanced robots where human oversight is still present and the final moral judgement is made by a human rather than the robot, attributing liability to the robot would not be feasible[375].

It has been argued that acknowledging the robot's own criminal liability could, in certain cases, lead to issues in the causal nexus between the

370 BECK, Über Sinn, 2013, p. 256.
371 NISSENBAUM, Accountability in a Computerized Society, 1996, p. 35.
372 BECK, Über Sinn, 2013, p. 256 f.
373 FREITAS/ANDRADE/NOVAIS, Criminal Liability of Autonomous Agents, 2014, p. 151; SWART, Constructing Electronic Liability, 2023, p. 594.
374 BECK, Intelligent Agents and Criminal Law, 2016, pp. 141-142.
375 HU, Robot Criminals, 2019, p. 512.

actions of the person behind the machine and the resulting harm[376]. In my view, however, rather than focusing on the legal recognition of the robot's own liability, the emphasis should be on assessing the extent to which the robot's conduct contributes to the harmful outcome and the corresponding reduction in human influence. In cases where the degree of influence is dominant, as discussed in the following section on causality, the question of whether the robot itself is held liable is irrelevant from the perspective of criminal law[377]. In criminal law, any individual who is at fault is held accountable. The fact that an entity assumes liability does not mean that others are absolved of it. Such an interpretation resembles the principle of shielding behind a corporate veil in private law or the search for a party liable for civil damages. However, the principle of fault in criminal law prevents this outcome. Hence, attributing liability to machines, which are inherently non-moral agents, while absolving the actual moral agent from accountability, leads to *scapegoating*[378].

On the other hand, it has been asserted that the absence of criminal liability for non-human entities in certain legal systems serves as a shield for offenders (in line with *societas delinquere non potest*). An example often cited is that, while an individual may face criminal liability for tax-related crimes, a company might not be held liable, allowing criminal liability to be circumvented[379]. There is a view that the recognition of robots' own criminal liability could have an indirect penalising effect on those who benefit from their use and thereby act as a deterrent. For instance, manufacturers would be incentivised to produce robots that do not cause harm, as they risk reputational damage if offences occur[380]. However, in my view, this argument is not compelling if the alternative of recognising robotic criminal liability is not a liability gap, but rather the accountability of the persons behind the machine. In such cases, holding these individuals liable would be more appropriate.

376 ALTUNÇ, Yapay Zekâ, 2021, p. 365.
377 See: Chapter 4, Section A: "Causality".
378 COOPER, et al., Accountability, 2022, p. 870; NISSENBAUM, Accountability in a Computerized Society, 1996, p. 34 f.
379 HALLEVY, Liability for Crimes Involving AI, 2015, p. 41.
380 HU, Robot Criminals, 2019, p. 509; Singapore, Report on Criminal Liability, 2021, p. 37 [para. 4.44].

b. Contra Arguments in Legal Literature Against AI-Personhood

The European Parliament's proposal to the Commission to grant AI-driven systems personhood under certain conditions was not aimed at conferring personal rights[381] on robots but rather addressing liability gaps by creating a target for civil liability claims[382]. However, it exceeded its initial aim and was subject to significant criticism. Indeed, the EU's High-Level Expert Group on AI did not support this proposal either[383]. In fact, a letter addressing the matter has been circulated for signatures (gathering 285 signatures as of 01.08.2025). The experts have argued that the notion of granting personal status to autonomous robots reflects an overestimation of current robotic capabilities, a misunderstanding of unpredictability and self-learning in robots, and is influenced by science fiction and sensationalist media coverage[384].

At their core, discussions surrounding electronic personality are fundamentally rooted in tort law and aim to create a "*sui generis* target for claims" through a "legal trick". While this may be an original idea for civil law, it is ineffective in criminal law, where the objective is not to ensure compensation for harm but to attribute fault and uphold justice[385]. Therefore, the essence of these discussions lies in the pragmatic need for creating subjects of civil liability.

The notion that AI systems could possess their own criminal liability is foreign to European legal culture and has found little support, because the criminal law framework has long been based on the individual culpable liability of natural persons. Even in legal systems that recognise derivative criminal liability for legal entities, an unlawful act is attributed to a specific natural person who, by virtue of their role or relationship, represents the legal entity[386]. In other words, corporations possess legal personhood

381 Whether machines could one day possess fundamental rights falls within the range of philosophy of law, not legal doctrine. See: HILGENDORF, Dilemma-Probleme, 2018, p. 678.

382 HILGENDORF, Dilemma-Probleme, 2018, p. 678.

383 High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 08.04.2019, https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (accessed on 01.08.2025); HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 550.

384 http://robotics-openletter.eu. (accessed on 01.08.2025).

385 MÜLLER, Roboter und Recht, 2014, p. 604; as cited in: SIMMLER/MARKWALDER, Guilty Robots?, 2019, pp. 19-20.

386 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 666.

because the relationships underpinning them can ultimately be traced back to human involvement[387]. Besides, even in this aspect, significant challenges persist. Subjecting a robot itself to financial compensation parallels long-standing arguments against holding corporations liable in similar ways. Traditional penalties, such as fines, often fall short in deterring corporate misconduct because they do not effectively hold the individuals within the corporation accountable. Instead, they may inadvertently harm unrelated innocent parties, including shareholders, employees, and consumers[388].

It is not reasonable to argue, at a factual level, that AI can be the subject of a crime based on its similarity to humans. The essence of criminal liability lies in being a moral agent, and for robots, this is not feasible in the foreseeable future. Even the most sophisticated robots cannot replicate human moral judgment because they lack the capacity to engage in essential moral reasoning processes. Even if robots could make decisions that appear indistinguishable from those made by humans, such decisions would still be morally deficient as they would not be made for the right reasons[389]. AI lacks free will because it is a system with predetermined objectives[390]. They do not possess the ability to comprehend their own autonomous structure, history, rights or obligations[391]. They cannot recognise the legal scope and content of their actions, control their behaviour and its social significance, or possess the capacity for responsibility. Moreover, they lack awareness of injustice and, as a result, lack the capability to bear punishment[392]. Particularly, they cannot be the subject of retributive punishment, as they are unable to comprehend its meaning[393]. Even if an AI-driven autonomous system may play a crucial role in the commission of a crime, it can only be a tool rather than autonomous agent, because deliberate intention is essential for moral agency[394].

It has been argued that it is unnecessary for the legislator to take the huge step of granting legal personality to complex autonomous systems to address liability gaps. Such gaps can be resolved without substantial issues,

387 VAN DEN HOVEN VAN GENDEREN, Do We Need Legal Personhood, 2018, p. 42.
388 COFFEE, No Soul to Damn, 1981, p. 389 ff., p. 407 ff.
389 PURVES/JENKINS/STRAWSER, Autonomous Machines, 2015, p. 851 f.
390 AKBULUT, Yapay Zeka, 2023, p. 307.
391 GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 416.
392 FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 877 f.; AKBULUT, Yapay Zeka, 2023, p. 308; ZHAO, Principle of Criminal Imputation, 2024, p. 33 ff.
393 GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 412.
394 COOPER, et al., Accountability, 2022, p. 870.

for instance, by extending strict liability[395], and deterrence can be achieved through other legal mechanisms[396].

Under no circumstances should a criminal law for autonomous systems lead to a premature exoneration of the persons behind the machine[397]. In fact, AI-driven autonomous systems cannot be regarded as responsible third parties whose intervention would exclude attribution, as their conduct does not constitute an action in the legal sense[398]. Therefore, no liability gap would arise[399]. Moreover, even in terms of civil liability claims, the creation of a liability fund may disincentivise the persons behind the machine, such as manufacturers or operators, to avoid harmful events as much as possible[400].

c. Synthesis and Evaluation

The question of whether AI-driven autonomous systems should be granted legal personhood has given rise to significant debate. To summarise the perspectives on this matter, proponents of this idea, some influenced by anthropomorphic perceptions, argue that advanced AI systems should be recognised as legal persons to address legal challenges such as liability gaps. They refer to examples such as the recognition of corporate personhood and other non-human entities as evidence to support their argument. Some emphasise the increasing complexity of AI and its capacity for human-like interactions, proposing that such systems, to address pragmatic needs, should be held accountable for damages, not merely as tools but as agents capable of assuming responsibility. On the other hand, the opposing viewpoint highlights that the absence of free will and moral agency (both of which are fundamental aspects of criminal liability) is a limitation inherent in AI. Even the most sophisticated AI is incapable of engaging in genuine moral reasoning or comprehending the consequences of its conducts, which precludes its suitability for criminal liability. European legal traditions, which are grounded in individual culpability, are reluctant

395 HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 551.
396 Singapore, Report on Criminal Liability, 2021, p. 5, [para. 25].
397 SCHUSTER, Künstliche Intelligenz, 2020, p. 393 f; FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 877 f.; IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 427 f.; TURNER, Regulating AI, 2019, p. 189 ff.
398 See: Chapter 3, Section B(3): "Can Autonomous Systems 'Act' In The Legal Sense?".
399 SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 505.
400 SCHUSTER, Künstliche Intelligenz, 2020, p. 393 f.

to extend personhood to non-human entities. Critics argue that existing mechanisms, such as strict liability of persons, can address accountability without altering the concept of personhood. They also express concern that attributing liability to AI may result in the evasion of liability by persons behind the machine, which would be inconsistent with the core principles of justice.

In one of the early discussions on the topic in 2007, the recognition of legal personhood for electronic agents was critically evaluated. While its necessity was acknowledged due to technological advancements, caution was advised regarding potential societal impacts and risks of alienation[401], which it could be argued, we experience today. According to one view, acknowledging such a category for robots would be like opening *Pandora's box*, leading to the recognition of personhood or expectation of free will and consciousness in other entities as well[402].

Attributing human-like characteristics to AI-driven autonomous systems frequently falls into the logical error known as the *android fallacy*. During 2024 and up to mid-2025, when this study was finalised, it was observed that society often responds with great enthusiasm to the remarkable achievements of AI, occasionally prompting the question with hype: has *Artificial General Intelligence (AGI)* finally arrived? However, while AI can perform tasks that are difficult for humans with relative ease, tasks that are simple for humans may still present significant challenges for AI. This situation cultivates anthropomorphic perspectives that align with the human evolutionary background, causing emotional biases to prevail over objective analysis and hindering the ability to assess reality as it is. For instance, if a robot equipped with software designed for voice communication is additionally fitted with actuators enabling facial expressions, people are prone to interacting with it as if it were human[403]. One day, a truly human-like or super-intelligence may indeed emerge (nothing is impossible), and

401 TEUBNER Gunther, "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law", Max Weber Lecture Series MwP 2007/04, 17.01.2007, https://hdl.handle.net/1814/6960, p. 20. (accessed on 01.08.2025).

402 LEHMAN-WILZIG, Frankenstein Unbound, 1981, p. 448.

403 A highly relevant phenomenon, which is named *Uncanny Valley*, describes the unsettling feeling that arises when a robot or humanoid figure closely resembles a human, yet exhibits subtle imperfections, leading to a significant decline in emotional affinity. First introduced by Masahiro Mori in 1970, this phenomenon occurs when near-human characteristics trigger discomfort due to perceptual mismatches or inconsistencies in appearance or behaviour. See: MORI, The Uncanny Valley, 2012, p. 98.

such developments may necessitate a renewed examination of these issues. However, it is more appropriate for scientific inquiry to be guided by the current evidence. Given the current state of technology, it is important to maintain a consistent approach, free from excessive influence by science fiction. It must be acknowledged that today's robots do not qualify as moral agents.

A recent study conducted by researchers from *Apple* and *DeepMind* demonstrates that LLMs lack true mathematical and logical reasoning capabilities, and instead rely on pattern-matching[404]. It was followed by another study conducted by *Apple*, which argues that despite notable improvements on reasoning benchmarks, current Large Reasoning Models (LRMs) still fail to demonstrate genuine reasoning capabilities or to comprehend in a manner comparable to human cognition[405]. This raises questions about our likelihood of achieving true reasoning as more advanced AI models are developed. It is true that AI technologies are rapidly advancing. For instance, an experiment with the earlier version of GPT (GPT-3) in 2020 involved providing the model with a text and asking it to complete the passage. GPT-3 often produced completions that were conceptually and logically absurd, such as suggesting attending a court hearing in a bathing suit[406]. However, subsequent versions of GPT have shown significant improvement, with reasoning that aligns more closely with human logic, indicating that AI is advancing swiftly and reaching more coherent and plausible conclusions. Yet, the question remains: will AI ever fully achieve human-like reasoning?

A debate persists among AI researchers: some contend that, given sufficient time and data, neural networks will eventually attain human-level intelligence. Others, however, dismiss this view as implausible at least for the foreseeable future. Admittedly, although I am still sceptical on this matter, the progress of AI from the beginning of this study to its submission as a doctoral thesis, and even up to the point of its submission for publication,

404 MIRZADEH Iman, et al., "GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models", arXiv, 07.10.2024, http://arxiv.org/abs/2410.05229, p. 1 ff. (accessed on 01.08.2025).

405 SHOJAEE et al., The Illusion of Thinking, 2025. However, the study has faced considerable criticism for potential bias, given that Apple had significantly lagged behind in the AI race as of mid-2025.

406 MARCUS Gary/DAVIS Ernest, "GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about", 22.08.2020, https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion. (accessed on 01.08.2025).

led me to reconsider my position (2020-2025). Although these systems do not, in any genuine sense, resemble human beings, externally observing their extraordinary ability to replicate human patterns of thought become more convincing. Indeed, the continuous evolution and increasing authenticity of the examples provided at the outset of this study (shifting with each update) indicate that this question will likely require reconsideration in the coming years. In other words, the illusion is so persuasive that it may soon become nearly impossible to distinguish it from genuine human patterns of thought.

Extensive *pro* and *contra* arguments concerning the possibility of AI becoming moral agents have been thoroughly analysed above. In my opinion, all arguments for recognising personhood in robots, apart from those based on pragmatic necessities, are inherently contradictory or misrepresent the essence of the concept. According to one view, against the shortcomings of current debates on the theoretical questions about the conditions necessary for moral agency and whether artificial entities can fulfil these conditions, we should focus on more practical and normative questions regarding how and to what extent they should be integrated into human social practices that traditionally involve moral agency and responsibility[407]. Another perspective presents that, for an entity to be considered a moral agent, it must possess traits such as rationality, free will, autonomy and phenomenal consciousness. To the contrary, functionalists maintain that moral agency is demonstrated through specific behaviours and responses, focusing more on external actions rather than the necessity of internal states[408]. Accordingly, it is noted that, given human consciousness is itself a subject of debate, consciousness should not be seen as an absolute prerequisite for personhood,[409] and conferring legal personhood does not necessarily require treating it as a human[410].

Adopting a pragmatic or functionalist approach to conferring personhood upon AI-driven systems would still present numerous inherent challenges. The foremost among these is the critical issue of determining how and to which entities personhood should be granted. This challenge stems from the fact that both non-physical and embodied systems can be easily created and distributed. Moreover, there is a wide range of AI-driven software: from internet cookies to sophisticated DNNs, which have been

407 BEHDADI/MUNTHE, A Normative Approach, 2020, p. 212.
408 For the assessment, see: *Ibid*, p. 198 f.
409 VAN DEN HOVEN VAN GENDEREN, Do We Need Legal Personhood, 2018, p. 41.
410 TURNER, Regulating AI, 2019, p. 205.

developed to perform a variety of functions. Under normal circumstances, an individual typically interacts with several distinct legal entities and dozens of natural persons daily. It is conceivable, however, that in the future, this number could expand to encompass thousands of interactions with different AI systems. Besides, in contrast to humans, software and hardware systems are not constituted of a single, unified entity. They can be divided, separated, integrated, combined, multiplied, disassembled and reassembled. By its very nature, recognising personhood for an entity 'emerging' from ones and zeros is inherently challenging. Moreover, when it comes to involvement in an offence, whether criminal or administrative; there is no distinction between being mere software and being a robot with physical hardware. Therefore, limiting personhood recognition solely to embodied systems is not a sufficient argument. If the criteria for granting personhood is registration or licensing, legal challenges may arise due to inconsistencies between the theoretical assumptions underpinning such registration and the practical realities. This divergence between legal expectations and real-world applications can lead to significant challenges.

Another issue is that legal entities conduct transactions through humans, ensuring human involvement in their operations[411]. In the case of AI, however, apart from a supervising individual or a designated human-in-the-loop, the person behind the machine -especially in the future- may not always be clearly involved or identifiable. As previously discussed, AI-driven systems can autonomously effect changes in the external world, much like viruses or bacteria, without direct human involvement.

Even if truly autonomous and intelligent robots come into existence in the future and are granted personhood, the associated person behind the machine may not be exempt from liability if conditions based on their own fault are met. Indeed, within the current criminal law framework, there is a general principle that individuals should not evade criminal liability by using robots as proxies for committing acts[412]. In response, it has been noted that an AI system is not entirely under human control and that its outputs may be unforeseeable. When a robot is intentionally used to commit a crime or cause harm, the person behind the machine would still be held accountable under existing laws[413]. However, in my opinion, the issue here is not about intentionally using or exploiting AI-driven autonomous

411 VAN DEN HOVEN VAN GENDEREN, Do We Need Legal Personhood, 2018, p. 42.
412 JOERDEN, Zur strafrechtlichen, 2020, p. 301.
413 TURNER, Regulating AI, 2019, p. 193.

systems, but rather about avoiding criminal liability risks, particularly in certain fields where such liability would typically arise through negligence (including in the context of civil liability). This could involve using AI systems, such as chatbots, and then scapegoating the system, relying on individual culpability, or non-attributability for the criminal result. Punishing robots is unimaginable in the foreseeable future[414]. Even if personhood were to be granted, it is ultimately humans who delegate tasks and endow robots with potentially unpredictable behaviour[415]. Therefore, liability for the machine's conduct should be attributed to those individuals, provided that the necessary conditions for fault are met[416].

Finally, for AI-driven autonomous systems to be considered criminally liable, they must first commit an act that constitutes an offence under criminal law. Only such an act could be the subject of examination under criminal law; if no act or omission exists from the perspective of criminal law, there is nothing to discuss[417]. Hence, any resulting harmful outcome will be attributed to the person behind the machine. The following section will explore whether robots can fulfil the *actus reus*. Ultimately, it will be concluded that they cannot, which results in the futility of any discussions on this matter from the outset.

3. Can Autonomous Systems 'Act' In the Legal Sense?

a. General Insights

As the level of autonomy in robots continues to advance, largely driven by an anthropomorphic perspective, expressions such as robots "killing", "injuring" or "saving" people have become more common in everyday language and are frequently mentioned in various news reports[418]. This

414 BECK, Die Diffusion, 2020, p. 45.

415 For a detailed discussion see: Chapter 4, Section C(5)(b)(3)(d): "Delegating Tasks to AI-Driven Autonomous Systems: An Alternative Approach for Liability".

416 NIDA-RÜMELIN/BAUER/STAUDACHER, Verantwortungsteilung, 2020, p. 94 f.

417 SEHER, Intelligent agents, 2016, p. 48.

418 "Robot kills worker at Volkswagen plant in Germany", 02.07.2015, https://www.theguardian.com/world/2015/jul/02/robot-kills-worker-at-volkswagen-plant-in-germany; SCHENEINER Bruce/OTTENHEIMER Davi, "Robots are Already Killing People", 06.09.2023, https://www.theatlantic.com/technology/archive/2023/09/robot-safety-standards-regulation-human-fatalities/675231; "Bear robot rescues wounded troops", 07.06.2007, http://news.bbc.co.uk/2/hi/health/6729745.stm;

linguistic framing frequently personifies them, ascribing human-like capabilities which in turn shape public perception of their capabilities. While robots equipped with physical embodiments are undoubtedly capable of effecting changes in and manipulating the physical world[419], the question remains whether their conduct can be considered as 'actions' in the context of criminal law. This issue hinges on the ability of robots to satisfy the criteria for *actus reus*, which traditionally necessitates the presence of a human actor capable of intentional conduct and possessing moral agency. Resolving this question is pivotal to determining whether robots can be classified as legally responsible agents.

According to the prevailing opinion and traditional doctrine in Germany, only natural persons can perform actions in the context of criminal law. Legal entities cannot act due to a lack of psychological and mental substance and they cannot express themselves. Instead, human agents can act on their behalf[420]. Therefore, even if robots could hypothetically be granted legal personhood, they would still not be considered capable of performing actions under this doctrine[421].

Various perspectives have been proposed on this matter, reflecting the differences between the common law and Continental European legal traditions. According to one view influenced by common law tradition, robots can fulfil both *actus reus* and *mens rea*. The movement of a robot's parts through mechanical or other mechanisms can be considered *actus reus*, and it is accepted that such behaviour can be attributed to the robot itself. Additionally, omissions can also be recognised; when a robot is under an obligation to act and fails to do so, its inaction can be regarded as an omis-

"Driver in fatal Tesla crash previously had posted video of autopilot saving him", 01.01.2016, 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); See also: GLESS, Mein Auto, 2016, p. 226.

419 CALO, Robotics and the Lessons, 2015, p. 530.

420 ROXIN/GRECO, § 8. Handlung in Strafrecht AT, 2020, p. 360 Rn. 59; HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 46 Rn. 11; RENGIER, § 7. Handlungslehren in Strafrecht AT, 2019, p. 42 Rn. 9; GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 141, Rn. 7; CORNELIUS, Künstliche Intelligenz, 2020, p. 61; ZIESCHANG, Strafrecht AT, 2023, p. 27 Rn. 48.
The discussion in Turkish law regarding the capacity of legal entities to act through their organs and the relevant legal norm was outlined above. See: Chapter 3, Section B(1): "Fundamentals".

421 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 174; ZHAO, Principle of Criminal Imputation, 2024, p. 36.

sion[422]. Another opinion argues that, similar to corporations with policies and goals that act intentionally, robots can engage in cognitive activities through neural networks and thus act intentionally[423].

An alternative viewpoint posits that it is challenging to assert that the conduct of AI-driven entities can be defined as actions at present. While acknowledging this, it can be argued that, even if the conventional physical component of action is overlooked, this conduct does not meet the criteria for acts due to the absence of both intentionality and social conformity. However, this may evolve as AI continues to advance[424]. Some even argue that if a programmer intentionally designs an AI system to cause harm, the robot itself would act as the direct agent carrying out the harmful behaviour, and therefore fulfilling *actus reus*[425]. In my opinion, an analogy can be drawn here using the example of employing an animal for the purpose of an attack. In such a scenario, it is not the animal's conduct that is examined, but rather the behaviour of the individual commanding or controlling the animal, that is assessed in the context of criminal law. Conversely, if an attack is carried out by a wild animal, such an incident cannot be regarded as an act within the framework of criminal law[426].

It should be noted that, considering bodily movements (e.g., the movement of mechanical parts) alone as the material element of an act is an outdated approach and would exclude intelligent agents composed solely of software. This perspective would overlook cases such as cybercrimes, where conduct like executing a *Denial of Service (DOS)* attack does not involve physical movement but still constitutes an offence[427].

b. Assessment Based on Theories of Action

According to the traditional approach, while it is not definitively established whether humans possess true free will (only the impression of such exists), criminal law requires that an act be carried out with it, implying

422 HU, Robot Criminals, 2019, p. 511; HALLEVY, The Criminal Liability, 2010, p. 187, 192.
423 HU, Robot Criminals, 2019, p. 520 f.
424 LIMA, Could AI, 2018, p. 682.
425 MÜSLÜM, Artificial Intelligence, 2023, p. 139.
426 ZIESCHANG, Strafrecht AT, 2023, p. 27 Rn. 48.
427 FREITAS/ANDRADE/NOVAIS, Criminal Liability of Autonomous Agents, 2014, p. 151; GLESS/WEIGEND, Intelligente Agenten, 2014, p. 571 fn.48.

the ability to refrain from committing the act and to choose an alternative course of action[428]. The capacity of robots to fulfil *actus reus* is rejected on the grounds that they cannot autonomously set goals for themselves and set out to achieve them[429]. However, there are various theories of action put forward in German legal doctrine, and it would be appropriate to briefly assess whether this issue leads to different conclusions according to these theories.

According to the *natural-causal theory* of action, crime is viewed as bodily movement driven by will, and actions beyond human control are excluded from consideration. The *final theory* posits that action is human behaviour directed by will towards a specific goal; meaning that it can only be deemed an action when interpreted in light of an intention. The *social theory* defines action as socially significant conduct that is controlled or controllable by will. The *personal theory of action* considers action as an expression of one's personality. Lastly, the concept of *intentional norm compliance capability* holds that an action is behaviour that could and should have been avoided by the offender to prevent the realization of a criminal offense, encompassing both active conduct and omissions, provided that the offender had the physical and intellectual capacity to do so[430].

Natural-causal theory: In the early 20th century, influenced by the natural sciences, criminal law sought to define human actions purely as physical processes driven by will, such as muscle movements or lack thereof. This concept has been criticised as being overly broad, potentially attributing a criminal outcome to anyone's actions or inactions, and is now considered outdated[431]. According to this theory, an action is a form of human behaviour that can be controlled by will (arbitrary act) and brings about a certain consequence in the external world[432].

Disregarding the prerequisite of being human, it has been argued that any "arbitrary bodily movement" could be considered an action from a purely external perspective, thereby permitting intelligent agents to be

428 JOERDEN, Strafrechtliche Perspektiven, 2013, p. 201, 203.
429 GLESS/SILVERMAN/WEIGEND, If Robots Cause Harm, 2016, p. 419.
430 KINDHÄUSER/ZIMMERMANN, § 5 Die Straftat als Normwiderspruch – Strafrecht AT, 2024, pp. 58-61 f. Rn. 10-21; GLESS/WEIGEND, Intelligente Agenten, 2014, p. 571 f.
431 STRATENWERTH/KUHLEN, § 6 Die Grundformen in Strafrecht AT, 2011., p. 56 Rn. 4 f.
432 JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 219; RENGIER, § 7. Handlungslehren in Strafrecht AT, 2019, p. 41 Rn. 3.

regarded as actors[433]. Accordingly, intelligent systems that evaluate data, develop, and make decisions, even in unpredictable ways, could be viewed as acting wilfully and thus having legal relevance. However, there is an ongoing debate about whether such systems genuinely act "wilfully" or merely follow pre-programmed, automated responses, leaving the question open to interpretation[434]. Thus, while the conduct of largely automated systems cannot be considered as acts under criminal law, that of autonomous systems which do not follow strictly predefined commands is open to question[435].

Final theory of action: As proposed by *Welzel*, an action is a purposeful human activity where individuals use their understanding of causality to foresee potential outcomes and anticipate a goal, select the means to achieve it, and consciously direct their behaviour to realise their will in the external world[436]. This notion outlines a rational structure of action, beginning with the conception of the goal, which is influenced by drives and interests, and continuing through the selection of suitable means and the weighing of side effects, to the decision and implementation[437].

According to the prevailing opinion, the conduct of AI-driven autonomous systems cannot be considered as actions under the final theory; because they cannot set their own goals and the system's decision-making power is merely derived from humans who developed the software and set the limits. Besides, despite their decision-making and autonomous learning capabilities, they lack wilful intent and an understanding of the social consequences of their conduct[438]. Some other scholars hold the same view, as they regard being human as a prerequisite[439].

Particularly, whether AI-driven autonomous systems make decisions based on predetermined programming or through their own evaluations is significant for future systems and remains a matter for external assessment. These systems cannot set themselves deliberate goals or direct their

433 GLESS/WEIGEND, Intelligente Agenten, 2014, p. 572.
434 WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 133 f.
435 REINBACHER, Social Bots, 2020, p. 462 f.
436 WELZEL, Das deutsche Strafrecht, 1969, p. 33.
437 JESCHECK/WEIGEND, Lehrbuch Des Strafrechts, 1996, p. 220 f.; STRATENWERTH/KUHLEN, § 6 Die Grundformen in Strafrecht AT, 2011., p. 57 Rn. 6 ff.; RENGIER, § 7. Handlungslehren in Strafrecht AT, 2019, p. 41 Rn. 4.
438 HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 29; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 134 f.; YUAN, Lernende Roboter, 2018, p. 481; QUARCK, Zur Strafbarkeit, 2020, p. 66 f.
439 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, pp. 259-260.

conduct toward these objectives. Achieving this would require the system to be aware of its conduct and understand its social significance, including its potential impact on others. At present, this level of recognition and perception in these systems is not considered possible[440].

It can be argued that the goal-oriented approach, which forms the basis of the final theory of action, exhibits similarities to a concept that is frequently utilised in AI development, particularly in regard to the identification of subtasks and the autonomous execution of them to solve a given problem. For more advanced future AI systems, the notion of goal-oriented conduct is indeed open to question. Nevertheless, it is unclear whether these systems are capable of acting beyond the objectives for which they were created[441]. In particular, final theory requires that behaviour must also be wilful, a quality that AI fundamentally lacks. Moreover, this theory was developed specifically to better understand and distinguish human behaviour from that which does not qualify as such. Therefore, attempting to apply it analogously to robots is not an appropriate approach; the same reasoning could be applied to intelligent animals, illustrating the limitations of such comparisons.

Social theory of action was initially developed to define legally relevant actions as functional social units of meaning. The theory was later expanded to encompass human behaviour as a response to situational demands using available options[442]. Accordingly, an action is any socially significant behaviour controlled or controllable by human will[443].

The personal concept of action is not fundamentally different from the social theory of action; in essence, it is a reflection of one's personality[444]. According to this theory, legal entities cannot express themselves as they lack psychological and mental substance; however, human agents can act on their behalf. Additionally, animals cannot act voluntarily or with purpose, and their actions do not qualify as "expressions of personality"[445]. Similarly, machines do not possess a personality to express, although the

440 GLESS/WEIGEND, Intelligente Agenten, 2014, p. 572, 578 f.
441 Here, to emphasise the autonomous nature of AI, the term programmed, which often evokes a deterministic if-then approach, has been deliberately avoided.
442 STRATENWERTH/KUHLEN, § 6 Die Grundformen in Strafrecht AT, 2011., p. 59 Rn. 12 f.
443 RENGIER, § 7. Handlungslehren in Strafrecht AT, 2019, p. 41 Rn. 5.
444 ROXIN/GRECO, § 8. Handlung in Strafrecht AT, 2020, p. 355 Rn. 44; STRATENWERTH/KUHLEN, § 6 Die Grundformen in Strafrecht AT, 2011, p. 59 Rn. 13
445 ROXIN/GRECO, § 8. Handlung in Strafrecht AT, 2020, p. 360 Rn. 58 f.

human operating or programming the machine does. Therefore, AI-driven autonomous systems cannot act in terms of criminal law[446].

In light of the aforementioned, it can be asserted that, within the context of criminal law, the notion of action -regardless of whether it creates a change in the external social world[447]- requires behaviour driven by will[448] and the capacity to understand norms as prerequisites[449]. However, based on current technology and empirical evidence, AI-driven autonomous systems are incapable of forming their own will and therefore cannot be considered capable of action[450]. It is argued that, perhaps only in the future, when a truly intelligent system capable of forming its own controllable will is developed, could it be considered capable of action in the sense of criminal law[451].

Additionally, legally relevant action is -regardless of any discussions about free will or determinism- limited to the behaviour of a person who can be directly addressed by the norms of the law, is capable of understanding these norms, and can reflectively incorporate this understanding into decisions regarding subsequent behaviour[452]. It is widely accepted that AI-driven autonomous systems lack this ability[453]. According to one view, these systems operate in accordance with the framework of pre-programmed norms that are implemented by humans and therefore do not fulfil the requirement of understanding norms[454]. In my opinion, however, it is not solely because AI-driven autonomous systems are pre-programmed by humans (or more accurately, trained and further developed using machine learning techniques) that they fail to meet the understanding of norms requirement. Rather, these systems cannot comprehend legal and social norms due to their inherent limitations. While the *code is law* approach is worth recalling, "understanding of norms" was not conceptualised to describe the algorithms of robots. The programming of these systems con-

446 *Ibid*, p. 369 f. Rn. 66f, 66g.
For the same view, see: IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 250.
447 v. LISZT, Lehrbuch des Deutschen Strafrechts, 1932, p. 154.
448 ZIESCHANG, Strafrecht AT, 2023, p. 27 Rn. 49 ff.
449 QUARCK, Zur Strafbarkeit, 2020, p. 66 f.
450 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 174.
451 WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 137; SCHULZ, Verantwortlichkeit, 2015, p. 95.
452 SEHER, Intelligent agents, 2016, pp. 48-50; QUARCK, Zur Strafbarkeit, 2020, p. 66 f.
453 SEHER, Intelligent agents, 2016, p. 51.
454 *Ibid*, p. 50.

sists solely of data and mathematical formulas, lacking the capacity for true comprehension of norms. Even if, in the future, these systems were to advance significantly, and Asimov's robotic laws[455] were somehow integrated into their software, it would not represent a genuine understanding of norms. Rather, it would most likely be an illusion of such understanding.

c. Re-interpretation of the Concept "Action"

It is evident that the prevailing doctrine on theories of action does not define the conduct of AI-driven autonomous systems as actions within the context of criminal law. This is either because being human is a prerequisite, or because robots are unable to fulfil the requirement of wilful behaviour, are not capable of understanding the norms, or lack the requisite personality to express. This phenomenon is understandable, given that criminal law was created by humans, for humans, and the direct application of these concepts to machines would be ineffective. Consequently, it is argued that rejecting from the very outset the application of concepts such as action, responsibility, and guilt -principles deeply embedded in human-centric jurisprudence- restricts the legal system's ability to address new challenges posed by autonomous systems[456]. Accordingly, if it is the desired outcome to recognise the liability of a robot, it may be necessary to set aside the requirement of wilfulness, as understood in the human sense[457]. Similarly for example, corporate criminal liability is recognised in many legal systems, where the act is not tied to the actions of the representative or individual organs, but to those of the company itself. In this way, an adaptation that aligns with the dynamics of new technology can be achieved[458].

In light of these explanations, *Hilgendorf* asserts that existing concepts can be reinterpreted over time to meet the needs of the era; concepts are not immutable in a linguistic sense. This is not a novel approach in the

455 Asimov's famous three robot laws were first introduced in the short story Runaround in 1942, and were later amended with a 'zeroth law'. ASIMOV, "Runaround", Astounding Science Fiction, Ed. John W. Campbell. New York: Street & Smith, 1942.

456 HILGENDORF, Können Roboter schuldhaft handeln?, 2012, p. 119 f.

457 OSMANI, The Complexity of Criminal Liability, 2020, p. 71 ff.; QUARCK, Zur Strafbarkeit, 2020, p. 67.

458 *Ibid.*

field of law. In fact, language is a living phenomenon, and the nuances of its conceptual content can undergo transformation and interpretation over time. Legal definitions also exert a formative influence on linguistic meanings, with the objective of achieving specific goals or addressing emerging necessities. In law, terminology is subject to a process of continuous evolution, driven by the need to adapt to shifting requirements through reinterpretation[459].

Robots can exhibit conduct in the sense of visibly recognisable bodily (mechanical) movements; however, it is difficult to assert that this conduct is controlled by will, and certainly that it does not resemble human will. It is possible to discuss the concept of "will control" in machines *via* their programming, through a different interpretation that considers this rule-based behaviour. *Hilgendorf* raises the debate on whether such behaviour by robots can be considered as actions. He acknowledges that this is a reductionist approach, neglecting the complexity of human volitional control and disregarding the contemporary scientific discussions surrounding the issue of free will. Yet, he draws an analogy by pointing out that, just as robots are programmed to behave in certain ways when specific conditions are met; humans also act in accordance with, or are guided by, certain rules. Thus, he opens up the discussion of reinterpreting robotic conduct through their programming as actions[460].

In response, reaffirming that such an analogy is reductionist and provides only a very incomplete perspective on the complexity of human volitional control; it has been argued that this approach represents a purely causal understanding of action, as it reduces the entire process to a series of if-then sequences[461]. Additionally, another view highlights the drawbacks of referring to both cases as actions. Accordingly, one could indeed redefine the concept of action so that, under this new definition, machines would also be considered capable of acting. However, such a reinterpretation would not be beneficial; instead, it would create the misleading impression that human action and machine action are identical phenomena[462].

In my opinion, acknowledging that language is a living phenomenon and that concepts evolve over time, the primary question that must be addressed is whether it is truly necessary to hold robots liable. Criminal law, along with its concepts and principles, was developed specifically

459 HILGENDORF, Können Roboter schuldhaft handeln?, 2012, pp. 122-124.
460 *Ibid*, p. 125 ff.
461 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 252.
462 ROXIN/GRECO, § 8. Handlung in Strafrecht AT, 2020, p. 370 Rn. 66 f.

for human beings. Therefore, applying these concepts to different entities through reinterpretation could lead to entirely new and complex problems. As elaborated above, granting personhood to robots is not possible *de lege lata*, and it is not needed *de lege ferenda*. Any justification for such a requirement can only be grounded in a pragmatic or functionalist approach, avoiding the pitfalls of the *android fallacy*. A similar rationale could apply to recognising robots as capable of performing actions; however, doing so would essentially be creating fictions in every aspect. If numerous legal fictions are to be established, one could equally apply this logic to categorise the flow of a river or the conduct of an intelligent animal as an action. Therefore, the key question is: is this required? At present, it can be argued that it is not. Should such a need arise in the future, we would require an entirely new legal framework rather than adapting or stretching our current legal institutions to accommodate these circumstances.

C. Various Liability Models for the Person Behind the Machine

Offences in which AI-driven autonomous systems are involved may concern not only criminal law, but also administrative and civil law. The inherent characteristics of criminal law presents a challenge in identifying the person behind the machine and their guilty act; and in some cases, such individuals may not be held criminally liable. To better highlight these challenges and to more precisely distinguish the points at which criminal liability diverges, it is essential to examine various other liability models. Through this analysis, it will be possible to assess whether these models can sufficiently contribute to the achievement of justice and, as suggested in literature[463], whether criminal law could benefit from these models to fill the contested "liability gap" in the future. For instance, in response to arguments advocating for the implementation of vicarious liability to the use of robots; it would be sensible to examine whether such scenarios truly stem from the actions of another party, as in employer-employee relationships. Adapting the established criteria and findings in this area to the context of robots could provide a more accurate basis for assessment.

Despite the existence of distinctive challenges, civil law liability models do not typically result in liability gaps. In certain situations, such as accidents involving self-driving vehicles, there may be an increase in cases

463 *E.g.:* ABBOTT/SARCH, Punishing Artificial Intelligence, 2024, p. 111 ff.

where no one is held criminally liable, but rather civil law liability in the form of compensation is pursued. Undoubtedly, some incidents may pertain solely to civil law without constituting a crime. The issue here, however, lies in the potential of impunity for actions traditionally performed by humans when they are delegated to AI-driven autonomous systems. This raises questions about the distinction between these two areas of law.

Addressing such violations solely through material remedies, such as monetary compensation or administrative fines, without subjecting anyone to criminal law sanctions, could undermine the functions of criminal law. Approaches that are becoming increasingly prevalent, particularly in Anglo-American law, which disregard the offender's culpability, would represent a paradigm shift and bring criminal law sanctions closer to administrative punishments[464]. However, a purely compensatory approach may fall short of meeting society's expectations for justice and may weaken the perceived legitimacy of the legal system. Humans are often driven by a retributive sense of justice and approaches which solely aim to deter future offences are insufficient[465]. The deployment of sanctions in other fields of law to address infringements may result in a retribution gap that can only be addressed through the mechanisms of criminal law[466]. Retributivism encompasses not merely the administration of deserved punishment, but also its moral necessity. From this perspective, retribution can be justified independently of utilitarian considerations, such as the consequences of the punishment[467].

With growing robotisation, it is inevitable that AI-driven autonomous systems will assume a more prominent role in the causal nexus of harmful outcomes. As previously discussed in detail, this may result in society attributing blame to robots as the perceived cause of harm; especially since evolutionary primitive instincts lead humans to express anger toward tangible objects. Nevertheless, robots are not suitable subjects for retributive blame, which creates a retribution gap[468]. Moreover, in the absence of punitive or pre-emptive measures, civil law remedies are inadequate, and even potential compensation fails to function as a real deterrent when absorbed

464 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 666.
465 JENSEN, Punishment and Spite, 2010, p. 2641, 2644; DANAHER, Robots, 2016, p. 299, 302.
466 DANAHER, Robots, 2016, p. 300 f.
467 MOORE, Justifying Retributivism, 1993, p. 21 ff.
468 DANAHER, Robots, 2016, p. 302, 308.

by industries or insurers that can incorporate them into their calculations in advance[469].

In a future where robots perform most tasks, it is pertinent to consider how the presence of a "retribution gap", rather than a "criminal liability gap", will affect society. Thus, from the standpoint of legal dogmatics and policy, the question becomes: in the event of a fatal multi-vehicle accident caused by a self-driving taxi, will the families of the deceased truly feel that "justice is served" by a sincere apology from the manufacturing company and compensation in the form of a five-figure sum in US dollars, when no one can be held criminally liable?

1. Can Civil Law Liability Models be Adapted to Criminal Law?

The fundamental objective of civil law is to achieve a fair and equitable distribution of social and economic risks through the allocation of financial burdens. In contrast, criminal law is primarily concerned with the utmost protection of legal interests and the rectification of breaches of fundamental societal norms. This is achieved through the imposition of blame and the assignment of severe sanctions, which are subject to stricter substantive and procedural standards due to the gravity of the penalties involved[470]. Civil liability operates on the principle of total reparation; meaning any injury, no matter its severity, qualifies for compensation. However, these principles cannot be directly applied to criminal law, which prioritises protecting individual freedom and social order rather than maximizing compensation for damages[471]. Hence, while civil law may recognise liability based on presumed fault or strict liability, criminal accusations apply only when there is proven faulty misconduct by an individual[472]. A proposed solution suggests that for offences involving AI-driven autonomous systems, the gaps in criminal liability and difficulties related to punishing the robot itself might be addressed by expanding civil liability and introducing targeted amendments to existing criminal law[473].

469 SCHUSTER, Künstliche Intelligenz, 2020, p. 389 f.
470 OEHLER, Die erlaubte Gefahrsetzung, 1961, p. 246; STUCKENBERG, Causation, 2014, p. 471; ASARO, A Body to Kick, 2012, p. 184; SAYRE, Criminal Responsibility, 1930, p. 721 ff.
471 BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 132.
472 GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 563.
473 ABBOTT/SARCH, Punishing Artificial Intelligence, 2024, p. 111 ff.

It is argued that the existing liability concepts in civil law provide instruments for an appropriate distribution of liability[474]. In addition to fault-based tort law provisions, the principles of strict liability, vicarious liability, and product liability may serve as valuable guides in determining responsibility and in facilitating harm correction and reduction[475]. Beyond these, models such as the liability of slave owners and the responsibility of animal keepers have also been frequently analogised in literature and will be briefly examined below. The analyses are purely theoretical. Thus, the issues that may arise in practice, such as burden of proof, pertain to real-world application and will not be detailed here.

In areas such as autonomous driving, liability issues that emerge within the context of criminal law can be more easily addressed through the utilisation of civil liability concepts, including those of strict liability and product liability[476]. Properly formulated liability rules enable producers and operators to exercise a legally adequate standard of care in the design, testing, monitoring, and operation of AI-driven systems[477]. In the absence of a robust and deterrent regulatory framework for AI, corporations engaged in AI development may not be sufficiently deterred from pursuing high-risk ventures, particularly in light of the considerable profit margins these entities have realised in recent years[478]. Under no circumstances, when an AI-driven system is implemented in place of a human to perform a task, should a liability structure be established that results in reduced accountability. The potential for liability should serve as an incentive for systems to be kept up-to-date and for greater caution to be exercised to ensure safe use. This is necessary to preserve a fair balance between benefit and burden. Additionally, those who suffer harm should not be provided with a more restricted right or opportunity for compensation[479].

In matters of civil law liability, the insurability of liability significantly facilitates the resolution of matters. Although a proposal has been made for a state accident insurance scheme that socialises the risks of robotics tech-

474 GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 573.
475 REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 75; YÜNLÜ, Current Developments on AI, 2019, p. 206.
476 HILGENDORF, Moderne Technik, 2015, p. 100; SCHUSTER, Das Dilemma-Problem, 2017, p. 102.
477 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 12.
478 MALGIERI/PASQUALE, Licensing High-Risk AI, 2024, p. 2.
479 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 12; YÜNLÜ, Current Developments on AI, 2019, p. 207.

nology and grants comprehensive exemptions from liability for producers and users; this approach neither promotes the safe development of robotics technology nor encourages its risk-free and careful use[480]. Furthermore, it has no impact on criminal liability. Conversely, criminal liability (unlike civil liability) cannot be mitigated by insurance or similar mechanisms[481].

The laws enacted in various countries regarding autonomous driving address matters related to registration, civil liability, and insurance, yet remain silent in addressing criminal liability matters[482]. Furthermore, several automotive companies have asserted their intention to assume liability for damages incurred while their vehicles are in autonomous driving mode[483]. While this declaration may not have direct implications from a criminal law standpoint, it could potentially be taken into consideration in the context of civil liability[484].

Liability disclaimers issued by companies, individuals, or institutions have no validity in criminal law. However, if such disclaimers thoroughly inform users of potential risks -such as when an AI-driven system is classified as experimental rather than a standard commercial product- or clearly state the possibility of malfunctions and the need for users to exercise utmost care, this may be considered as obtaining informed consent or other legal mechanisms[485]. In civil law, particularly under Turkish law, clauses disclaiming liability for gross negligence are absolutely void, though disclaimers for slight negligence may be enforceable.

a. Fault-Based Torts Liability

In the context of civil law, fault-based torts refer to wrongdoings entailing liability for damages resulting from "faulty conduct" (intentionally or neg-

480 ZECH, Zivilrechtliche Haftung, 2016, p. 202.
481 BECK, Selbstfahrende Kraftfahrzeuge, 2020, p. 446 Rn. 29.
482 THOMMEN, Strafrechtliche Verantwortlichkeit, 2018, p. 27.
483 GOLSON Daniel, "We put our blind faith in Mercedes-Benz's first-of-its-kind autonomous Drive Pilot feature", 27.09.2023, https://www.theverge.com/2023/9/27/23892154/mercedes-benz-drive-pilot-autonomous-level-3-test; KOROSEC Kirsten, "Volvo CEO: We will accept all liability when our cars are in autonomous mode", 07.10.2015, https://fortune.com/2015/10/07/volvo-liability-self-driving-cars. (accessed on 01.08.2025).
484 See, for similar views: DOĞAN, Sürücüsüz Araçlar, 2019, p. 3245.
485 For a wide assessment of consent under Turkish law from a medical law perspective, see: GÜVENÇ, Aşı Karşıtı Veli, 2022, pp. 32-47.

ligently) that infringes upon protected rights or interests[486]. For example, if a (so-called) "robot employee" causes harm, non-contractual liability arises for the individual or entity behind the machine under Section 823(1) of the German Civil Code (BGB) or Article 49 of the Turkish Code of Obligations[487]. Such liability requires unlawful and culpable conduct that infringes on specified rights or interests (such as life, person, health, freedom, property, or other protected rights) through actions causally linked to the resulting harm[488]. For instance, if a self-driving taxi causes an accident by hitting a pedestrian while transporting a passenger, the issue falls under tort liability for the pedestrian. Conversely, with respect to the passenger, contractual liability arises[489]. In this regard, the autonomous nature of the taxi is irrelevant[490].

In contrast to strict product liability, which will be examined below, producer liability under Section 823(1) is based on the principle of fault in line with general tort law principles. The liability of a manufacturer for harm caused to third parties due to a defective product exemplifies a classic case of tort liability. For this type of liability to arise, the harmful act must be unlawful and culpable, and there must be a causal link between the act and the resulting harm[491].

Fault-based liability incentivises individuals to interact with the system in greater caution and diligence, ensuring adherence to their responsibilities and the standard of due care. However, this presupposes that the prerequisites for the permitted use of technology, *i.e.* the specific duties of care, are clearly recognisable[492]. Classical tort law is fundamentally based on the principle of foreseeability. This concept entails a type of predictable harm affecting a foreseeable group of potential victims[493]. In the context of AI-driven autonomous systems; manufacturers, programmers, and sellers as well as the operators of these systems may be held liable if they could have reasonably foreseen or implicitly accepted that the machine's use might result in material or bodily harm. However, determining liability

486 MARKESINIS, German Law of Torts, 2019, p. 15.
487 PAGALLO, The Laws of Robots, 2013, p. 115; YÜNLÜ, Current Developments on AI, 2019, p. 199.
488 MARKESINIS, German Law of Torts, 2019, p. 29.
489 Yet, the existence of a contractual relationship does not preclude tort liability.
490 YÜNLÜ, Current Developments on AI, 2019, p. 201.
491 FUCHS/BAUMGÄRTNER, Ansprüche aus Produzentenhaftung, 2011, p. 1058.
492 ZECH, Zivilrechtliche Haftung, 2016, p. 197.
493 KARNOW, The application, 2016, p. 72.

becomes more complex when damage arises from systems operating as intended but encountering unforeseen, exceptional circumstances[494].

Criminal offences, while sharing certain similarities with fault-based torts, diverge significantly in several key aspects. In essence, the objective of torts is to compensate the injured party, whereas criminal law is primarily concerned with punishment as a means of retribution, deterrence and the prevention of recidivism[495]. Additionally, negligent liability in criminal law is exceptional and must be explicitly prescribed by statute, unlike in tort law. Moreover, multiple perpetrators in criminal cases are punished separately, according to their individual acts and degrees of guilt. In contrast, in tort law, a single amount of compensation is determined and paid either jointly and severally or according to each individual's share of responsibility. Furthermore, due to the *nulla poena sine culpa* principle, strict liability is not admitted in criminal law, whereas this does not apply in tort law. Criminal liability is personal, while in tort law, as will be discussed below, liability for another's actions (*vicarious liability*) is possible. Furthermore, although counterexamples can be provided, in principle, every crime constitutes a tort, but not every tort constitutes a crime[496].

In one of the earliest rulings concerning technological assistance systems, the Munich District Court (*Amtsgerichts München*) held in its judgment of 2007 that a driver was liable for damages when the parking assistance system failed to signal due to a hollow space. The court highlighted that drivers must not solely rely on such technology and must also ensure safety through their own observation[497]. This decision emphasises a fundamental yet pivotal point regarding the future of AI-human interactions: the vital importance of the supervisory role when the ultimate decision-maker is human. However, there are cases where the entirety of a task may be delegated to an autonomous system. Even in such instances, the supervisory

494 HILGENDORF, Recht und autonome Maschinen, 2015, p. 15.
Instead of liability, a model has been proposed in which insurance directly compensates the victim of an accident for damages. However, this approach has been criticised for lacking a deterrent effect. See: LOHMANN, Liability Issues, 2016, p. 339

495 ZIMMERMANN, The Law of Obligations, 1992, p. 902.

496 ZIMMERMANN, The Law of Obligations, 1992, p. 902; BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 134.

497 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.

role of humans remains significant. This matter will be discussed in detail throughout the study.

A significant point for consideration is the extent to which civil law standards can be made use of in defining the limits of negligent liability in criminal law. The duties of care in civil law and criminal law are not always congruent, as civil law pursues different objectives than criminal law, namely the balancing of property interests. Because of the insurability of risks, civil law standards of due care can be set higher than those in criminal law[498] and these standards set the upper limit for criminal liability. Therefore, not every instance of fault-based tort liability necessarily entails criminal liability; however, in cases where tort liability cannot be established, criminal liability should also be rejected[499]. In tort law, technical standards play a significant role in determining the objectively required standard of care, even if they are not legally binding on the court[500]. While such standards are also significant in criminal law, as will be discussed below[501], relying on them to determine the standard of care in criminal liability can raise concerns[502]. Another critique concerns the tendency to emphasise the differing goals and rules of tort and criminal law without engaging in a substantive debate on the matter. In this context, the German Federal Court of Justice's (BGH) *Lederspray* decision of 1990[503] is crucial, as it warned against using civil law principles to decide criminal cases without careful consideration[504]. The definition of negligence in civil law (Section 276(II) of BGB)[505] only emphasises failure to exercise the care required by ordinary and is unsuitable for criminal law because civil

498 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).

499 BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 134.

500 ZECH, Zivilrechtliche Haftung, 2016, p. 183.

501 See: Chapter 4, Section C(5)(c): "The Feasibility of Defining Permissible Risk Through Standards and Other Norms of Conduct".

502 BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, p. 133.
See also: VALERIUS, Sorgfaltspflichten, 2017, p. 21.

503 Federal Court of Justice (BGH), judgment of 06.07.1990, Case No. 2 StR 549/89, (*Lederspray case*), reported in NJW 1990, p. 2562.

504 BLECHSCHMITT, Der Fahrlässigkeitsmaßstab, 2015, pp. 131-132.

505 Bürgerliches Gesetzbuch (BGB), enacted on 18.08.1896, last amended on 23.10.2024. § 276 Verantwortlichkeit des Schuldners: "(2) Fahrlässig handelt, wer die im Verkehr erforderliche Sorgfalt außer Acht lässt." https://www.gesetze-im-internet.de/bgb/__276.html. (accessed on 01.08.2025).

law focuses on compensating for damages, whereas criminal law aims to condemn personal misconduct, making their objectives and consequences fundamentally different[506].

b. Vicarious Liability

(1) Respondeat Superior

Given the autonomous nature of AI-driven systems and the special relationship between certain parties or entities, it has been proposed that the *respondeat superior* model could be applied, drawing parallels with approaches from Ancient Rome. This analogy is based on the idea that, similar to the Roman legal principles that imposed liability on masters for the actions of their slaves or dependents; modern legal frameworks could extend vicarious liability to those who have a controlling or supervisory relationship over an autonomous system. Accordingly, damages caused by "robots" should be compensated by their owners or developers[507]. In a scenario where an AI system is not recognised as an agent[508], vicarious liability could only apply between the manufacturer (employer) and the programmer (employee). Conversely, if an AI-driven autonomous system is considered an agent, vicarious liability might also be applicable where it functions as an agent contributing to the outcome, thereby forming part of the relationship. There are differing views regarding who should bear vicarious liability for the AI in such cases. Some argue that AI systems, as agents, should give rise to vicarious liability for the owner or user[509]; while others contend that the manufacturer should be held responsible[510].

Vicarious liability, originating from the doctrine of *respondeat superior*, initially assumed that employers had control over their employees and were liable for their misconduct. Over time, this concept has evolved, extending beyond the employer-employee relationship and adapting to modern work structures like independent contractors; focusing on protecting victims

506 FREUND, § 5 Das Fahrlässigkeitsdelikt, 2009, p. 164 Rn. 13.
507 ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 385.
508 By "agent", reference is not made to the "AI agents" that became a subject of hype in 2025.
509 ASARO, A Body to Kick, 2012, p. 176 ff.; TURNER, Regulating AI, 2019, p. 101 ff.
510 GLAVANIČOVÁ/PASCUCCI, Vicarious Liability, 2022, p. 28.

rather than employer control[511]. In this context, various models of vicarious liability -such as those for children, employees, servants, slaves, and agents- have been suggested to address liability[512].

Studies examining the historical background of the *respondeat superior* doctrine since the 13th century indicate that the concept has undergone significant evolution and was not historically applied in the same manner as it is known today[513]. Between the 13th and 17th centuries, the doctrine applied solely in cases where the master had specifically commanded or authorised the servant to commit a tortious act or had provided consent before, or approval after, its commission[514]. In 1765, *Blackstone* described it as applying "if done by his command, either expressly given, or implied"[515]. By the 19th century, *respondeat superior* had taken on its modern form, where the notion of an "implied command" was replaced by the concepts of "course of business" and "scope of employment." This transformation led to an aspect of strict liability, under which the master could not escape liability, even if the act was contrary to an express command[516]. In other words, throughout history, this doctrine has been applied in the context of fulfilling a superior's command rather than examining *detour and frolic*. However, in the case of AI-driven autonomous systems, clear commands lead to intentional torts or crimes, which do not present issues. The challenge arises when autonomous systems cause harm which is either related or unrelated to the performing of the assigned task (*detour and frolic*).

The concept of vicarious liability presupposes that AI-driven systems are characterised as agents, whereas negligent liability and product liability regard them as objects[517]. Although one opinion suggests that AI must be regarded as a "tool" for vicarious liability to apply[518], having a certain degree of autonomy is more appropriate for the modern understanding and

511 *Ibid.*
512 TURNER, Regulating AI, 2019, p. 101.
513 For example, in the Statute of Westminster II of 1285, the phrase was used to denote the statutory liability of a public official for the misconduct of a subordinate in the performance of public duties, but only if the subordinate was unable or unwilling to pay for their own wrongdoing. See: SAYRE, Criminal Responsibility, 1930, p. 690.
514 *Ibid*, p. 691 f.
515 For the information, see: *Ibid*, p. 693.
516 *Ibid.*
517 TURNER, Regulating AI, 2019, p. 101.
518 ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 387.

application of this concept[519]. Therefore, if this model is to be applied, the first requirement is the categorisation of AI as agent, rather than a tool.

Vicarious liability of the superior is justified in the idea of control and benefit[520]. To illustrate, in the event of a waiter spilling wine on a customer, this is a foreseeable and potentially damaging occurrence within the context of business and the employer, who profits from it, should bear the responsibility. Accordingly, it is noted that, since robots are generally used for narrowly defined tasks, such as lawn mowing, this model could be applied[521]. As another example, if a robot is used for patrol duty by the police, even if the police did not manufacture the robot themselves and did not permit or intend an assault, liability may arise if the assault occurred within the scope of the robot's assigned role[522]. However, as the autonomy and purpose of robots increase, applying this doctrine will become increasingly difficult[523]. Indeed, not all activities of AI-driven systems can be encompassed, nor can all be attributable to the person behind them. The further AI strays from its delineated tasks, the greater the likelihood of a gap in liability arising[524].

In a recent case where *Air Canada*'s online chatbot provided misleading information that resulted in financial loss to a customer, the company argued that the chatbot constituted a separate legal entity and is responsible for its own conduct. Discussing the claim for negligent misrepresentation, the tribunal correctly stated that, "[w]hile a chatbot has an interactive component, it is still just a part of Air Canada's website. It should be obvious to Air Canada that it is responsible for all the information on its website. It makes no difference whether the information comes from a static page or a chatbot"[525].

519 LEHMAN-WILZIG, Frankenstein Unbound, 1981, p. 452.
520 JANAL, Die deliktische Haftung, 2016, p. 161.
521 REVOLIDIS/DAHI, The Peculiar Case, 2018, pp. 67-68.
522 TURNER, Regulating AI, 2019, pp. 100-101.
523 REVOLIDIS/DAHI, The Peculiar Case, 2018, pp. 67-68.
524 TURNER, Regulating AI, 2019, p. 101.
525 Moffatt v. Air Canada, 2024 BCCRT 149 (CanLII), 14.02.2024, https://canlii.ca/t/k2spq. (accessed on 01.08.2025).

(2) Exploring Existing Frameworks: Slavery, Animal Ownership, Employees and Associates

The application of *respondeat superior* to AI-driven autonomous systems is often compared to existing and historically applied models. One such comparison is the *noxal liability* and the status of slaves in Ancient Rome; where slaves, despite possessing equal intellectual capabilities to their enslavers, were regarded as property without rights or obligations. As a principle, they lacked the legal capacity to enter into binding agreements on their own. However, harm could still be caused to non-slaves both during the course of their duties and outside the scope of those duties[526].

Noxal liability describes the responsibility of a master for the actions of their slaves or a father for the actions of their children. Under this principle, if a slave or child committed harm or theft, the master or father could either give compensation for the damage or surrender the individual responsible (the slave or child) to the aggrieved party as a form of restitution[527].

Applying the master-slave analogy to AI-driven systems is challenging, considering the status of slaves in Ancient Rome was highly complex and evolved over time. Moreover, certain merits or values that could be considered akin to rights were eventually recognised to slaves[528]. Furthermore, while it is argued that AI should not be assigned the status of a slave, as slavery is a primitive concept that should be abandoned[529], it could be argued that it is more constructive to approach the matter analytically rather than dogmatically. At each stage, the reasons for such a stance should be examined considering the development of AI; particularly from the perspective of its possibility to attain a synthetic consciousness. This is because, at present, human interaction with even highly advanced computer systems and all inanimate objects is fundamentally based on absolutely exploiting them.

Despite their autonomous nature, slaves were legally classified as things, though they had a certain degree of legal recognition. Animals, on the other hand, possess autonomy of a different kind and lack legal personhood. Hence, comparison to trained animals provides a more compelling analogy for evaluating the potential or appropriate legal treatment of AI-driven au-

526 REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 69.
527 BUCKLAND, The Roman Law of Slavery, 1970, p. 98; REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 69.
528 BUCKLAND, The Roman Law of Slavery, 1970, p. 2.
529 BAK, Medeni Hukuk, 2018, p. 219.

tonomous systems[530]. Indeed, although owners maintain a degree of control over them, animals can act autonomously and sometimes in unpredictable, dangerous, and unexpected ways[531]. Similar to AI, they should be trained not to cause harm, and if they are deliberately trained to be vicious, commanded to attack, or inadequately restrained, the owner's liability may arise based on negligence or even intent. In such cases, the discussion focuses on the owner's mental state and intention rather than that of the animal[532]. The opposing view on the other hand, argues that equating AI to animals is unjustified, as AI's operations are based on algorithmic processes that resemble human rationality, with only limited parallels to the instinctual and sensory capacities of animals[533].

In German law, liability for animal ownership distinguishes luxury animals (such as pets) from animals domesticated for the purpose of enhancing the economic well-being of their owners. Strict liability applies to luxury animals, whereas for economically valuable animals, an owner can evade liability by proving either that an appropriate standard of care was exercised or that the harm would have occurred even if due care had been applied in accordance with Section 833 of the BGB[534].

Resembling AI-driven autonomous systems to employees or associates and their relationship with their superior; the question arises whether the autonomous system's conduct can be attributed to the operator within the context of vicarious liability when the operator's direct liability cannot be determined[535]. In the context of tort law, it refers to whether the employer can be held liable for the wrongful acts of an employee provided that these acts occur within the course of employment[536].

In German law, liability in such relationships is structured based on presumed fault. According to Section 831(1) of the BGB, a principal is liable for the unlawful and negligent conduct of their vicarious agent unless they can demonstrate that they exercised due care in selecting, managing and supervising the agent; or that the damage would have occurred even if

530 REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 70; SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 416.
531 ZECH, Zivilrechtliche Haftung, 2016, pp. 195-196.
532 ASARO, A Body to Kick, 2012, pp. 176-177.
533 ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 386.
534 REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 70; ZECH, Zivilrechtliche Haftung, 2016, p. 195 f.
535 SCHULZ, Verantwortlichkeit, 2015, p. 138; JANAL, Die deliktische Haftung, 2016, p. 150 ff.
536 MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 135.

the vicarious agent had been carefully selected and supervised[537]. Similarly, in accordance with Article 66 of the Turkish Code of Obligations, the employer is obliged to compensate for the damage caused to others by the employee during the performance of the work assigned to them. It has been argued that, since there is no provision stipulating that the term "employee" must exclusively refer to humans, the term in the provision can be broadly interpreted to include AI-driven systems as well[538]; yet this opinion is open to criticism from multiple perspectives.

In the context of vicarious liability of associates, an individual does not need to be capable of culpability to be considered an associate. It is possible to regard even a person lacking discernment as such. However, AI-driven autonomous systems cannot be classified as associates within this framework and liability for damages caused by them cannot be assessed under this rule (as per Section 278 of the BGB or Article 116 of the Turkish Code of Obligations). Consequently, these systems can only be considered extensions of the individuals utilising them[539]. Nevertheless, if AI-driven systems are granted legal personhood in the future, it may become possible to discuss the liability of the human employer or liability for associates in this context[540].

The adoption of a regulatory model for AI-caused liability, similar to occupational health and safety legislation has been proposed by the *Singapore Academy of Law Reform Committee*. According to this approach, certain designated units are required to implement all reasonably practicable measures to prevent harm. In workplaces, duties are assigned to occupiers and employers. Similarly, for AI systems, responsibilities could be allocated to entities best positioned -based on their proximity to and control over the system, as well as their resources- to take preventive, corrective, and mitigative actions against risks posed by AI and to shape future outcomes. This proposal advocates a shift from a broad, undefined liability framework to a more targeted, responsibility-based model, which is crucial for establishing clear legal expectations in the dynamic field of AI technologies[541].

537 JANAL, Die deliktische Haftung, 2016, pp. 151-152.
See also: ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 385.
538 SELANIK, Adam Çalıştıran, 2022, p. 358.
539 SCHULZ, Verantwortlichkeit, 2015, p. 138 ff.; YÜNLÜ, Current Developments on AI, 2019, p. 198 f.
540 YÜNLÜ, Current Developments on AI, 2019, p. 199 f.
541 Singapore, Report on Criminal Liability, 2021, p. 41, [para. 4.58 ff.].

(3) Applying Vicarious Liability in Criminal Law

Some scholars argue that vicarious liability may have a (limited scope of) application within criminal law. Accordingly, it has been proposed that cases involving the unpredictability of AI systems' outputs, which are examined under the category of negligent crimes in Continental European legal tradition are examined through the application of the legal concept of *respondeat superior* in the Anglo-American legal environment[542]. A further viewpoint posits that, since the primary aim of criminal law is to ensure deterrence, from a legal policy perspective, it may be considered acceptable to hold the master (superior) liable for certain minor offences ("petty misdemeanours involving no moral delinquency" in common law systems)[543] committed by a servant, even if these offences are unauthorised or unknown to the master. However, the *respondeat superior* doctrine should not be extended to cover serious or "true crimes" within criminal law, as this would misalign with the principles of personal culpability and proportionality inherent to criminal justice[544].

In the context of employment relationships, whether principals are obligated to prevent work-related offences committed by others (such as employees) and thereby incur criminal liability is a subject of considerable debate. It has been argued that such a guarantor position may be applicable only in the case of inherently dangerous enterprises[545]. Roles and positions assumed by individuals within legal entities, such as serving as an employer in a corporate structure do not by themselves, constitute a source of liability or responsibility under criminal law (the guarantor duties should be evaluated separately). This is because liability arising solely from a position reflects a strict liability approach, which may only be applicable in civil law. For criminal liability in negligence, the violation of a duty is a necessary precondition; however, this alone is insufficient. The breach of the relevant duty may not, by itself, significantly increase the risk of the occurrence of the harmful outcome[546].

542 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 665.
543 In U.S. law, *petty misdemeanors* are minor offenses, often not classified as 'crimes' in a strict sense, typically punishable by fines rather than imprisonment. See: REINBACHER, Das Strafrechtssystem der USA, 2010, p. 28, 142.
544 SAYRE, Criminal Responsibility, 1930, p. 722.
545 ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 178.
546 İÇER, İş Kazaları, 2020, p. 19.

In legal systems which recognise corporate criminal liability (like Swiss law), employees' actions are not directly attributed to the company without an independent accusation of organisational fault. However, if a programmer's negligence stems from inadequate infrastructure or control mechanisms within the company, the company may be accused of failing to do everything possible to minimise such errors[547].

To sum up, the advent of increasingly complex industrial processes since the 18th century has transformed the performance of tasks from individual efforts to collaborative operations facilitated by horizontal and vertical work relationships. Subsequently, delegation of many tasks to agents and subordinates has necessitated the accountability of a responsible superior under civil liability principles, a trend that is expected to continue with the integration of future technologies and autonomous systems. While this approach offers practical solutions in civil law, in criminal law, attempting to apply *respondeat superior* for another's actions, by disregarding the core principles of criminal law, raises concerns[548].

For instance, in vicarious liability under private law, the focus is not on the *mens rea* of the principal but rather on the relationship between the principal and the agent. In contrast, although there are differing opinions on the matter, the *mens rea* of the principal plays a pivotal role in criminal law[549]. The aim of criminal law is to protect social interests; in contrast to civil liability, which primarily seeks to identify a party responsible for compensating harm[550]. Therefore, vicarious liability conflicts with the foundational principles of causality and individual culpable liability in criminal law. Causation can only be established through "authorisation, procurement, incitation or moral encouragement" or the "knowledge and acquiescence" of a person[551]. Such a liability can only be feasible when the law explicitly departs from the general principles of criminal law (for example, by expressly penalising a crime committed by an agent in the course of its master's business)[552]. Therefore, the concept of vicarious liability is fundamentally incompatible with the principles of criminal law, as it

547 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 178.

548 SAYRE, Criminal Responsibility, 1930, pp. 716-717.
In a significant case from 1730, a judge who played an important role in the conceptualisation of the modern *respondeat superior* doctrine stated definitively that this doctrine should not apply in criminal law. See: *Ibid*, p. 701

549 TURNER, Regulating AI, 2019, p. 119; SAYRE, Criminal Responsibility, 1930, p. 721.

550 SAYRE, Criminal Responsibility, 1930, p. 721 ff.

551 *Ibid*, p. 702.

552 SAYRE, Criminal Responsibility, 1930, p. 712.

undermines the core notions of causality and personal culpability. Within this framework, attributing liability based on another's intent or state of mind could be considered inconsistent with these foundational principles.

c. Strict Liability

(1) Strict Liability Over Fault-Based Liability

The application of fault-based liability is often impeded by the complexities associated with establishing the foreseeability of an incident and proving causation. On the other hand, the concept of no-fault liability offers potential solutions (or shortcuts) to these challenges[553]. For instance, in the *Aschaffenburg case* described above[554], where the driver suffered a heart attack and lost consciousness, but the vehicle continued moving due to its lane-keeping system, resulting in death and injury, there is no issue regarding civil liability under Section 7 of the German Road Traffic Act (StVG). In such a fatal accident, which occurred during the operation of the vehicle, the owner is obliged to compensate the injured party for the resulting damage. This constitutes a form of strict liability, which can only be avoided by proving force majeure. In contrast, determining fault-based liability in this case is challenging. Neither the owner nor the driver could have foreseen the heart attack[555]. While it might be argued that the manufacturer should have anticipated this general possibility and taken preventive measures, it is difficult to reach a definitive conclusion on this matter given that, in 2012, the technology and general experience were still in its infancy. However, it is now evident that the proper technology must be implemented.

Historically, the concept of liability was rooted in pure causation. The shift to fault-based liability, specifically tied to negligence, represents a later development in legal thought[556]. While fault-based liability has become the predominant model, the transformative changes brought about by the Industrial Revolution with its transformative advancements, necessitated the adoption of strict liability as an exceptional legal mechanism to balance

553 TURNER, Regulating AI, 2019, p. 104; ASARO, A Body to Kick, 2012, p. 173.
554 See: Chapter 2, Section C: "Prominent Cases Highlighting AI-Related Liability".
555 HILGENDORF, Autonome Systeme, 2018, p. 104; HILGENDORF, Automatisiertes Fahren und Recht, 2018, p. 802 ff; HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 555.
556 KARNOW, The application, 2016, p. 63.

the societal benefits and risks brought about by new technologies[557]. It is further argued that the future adherence to the principle of fault as an absolute basis for liability remains uncertain, given that it already leads to unjust outcomes in certain cases today[558].

In the context of a fault-based liability framework, it is necessary to prove the culpability of the tortfeasor, the occurrence of harm or disadvantage, and the existence of a causal connection between them. By contrast, demonstration of the occurrence of harm or the risks posed by the wrongdoer, without the need to prove their intention or negligence, simplifies the legal process. Such a model can be justified not only in cases involving the control of animals or children but also for harm caused by AI-driven autonomous systems[559]. Indeed, even in simple computer programmes, bugs and harmful outcomes can occur despite all precautions. In this regard, strict liability is considered an effective solution for compensation in cases involving mass-produced products. It not only protects society by encouraging manufacturers to reduce risks but also ensures that victims can seek redress from the party best equipped to bear the cost. Additionally, it eliminates the significant challenges and economic burdens associated with proving fault, which can often be exceedingly difficult[560].

Considering these challenges in fault-based liability, applying strict liability not only in civil law but also in criminal offences involving AI-driven autonomous systems has been proposed[561]. Indeed, in such offences, the inability to identify a liable party under fault-based liability models may result in the harm being considered as mere "bad luck", leaving the victim and society to bear the burden, thereby creating a liability gap. Strict liability represents an effective policy for the prevention of such outcomes, as it provides an incentive for manufacturers to produce systems with lower risks and ensures that liability is attributed to those best positioned to implement

557 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 445.

558 VOGEL Joachim, BÜLTE Jens, Vorbemerkungen zu den §§ 15 ff, Strafgesetzbuch: Leipziger Kommentar: Grosskommentar, 13. Auflage, Band 1, CIRENER Gabriele, et. al. (eds.), Berlin: De Gruyter, 2020, p. 1022, Rn. 21.

559 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 3. For the argument that strict liability will remain functional for AI-driven systems until they become fully autonomous, then the point of focus should shift to AI's own responsibility, see: BAK, Medeni Hukuk, 2018, p. 225.

560 NISSENBAUM, Accountability in a Computerized Society, 1996, p. 39.

561 ZHAO, Principle of Criminal Imputation, 2024, p. 26 f.

preventive measures[562]. It has also been argued that, under Swiss law for instance, the concept of guilt has already diminished in absolute significance within administrative and corporate criminal law, being replaced by a more pragmatic equivalent[563].

The proponents of applying strict liability in criminal law put forward several arguments to support their position. First, they contend that strict liability encourages individuals to exercise greater caution, thereby raising overall standards of conduct. Second, they argue that it promotes procedural efficiency during the adjudication process by simplifying the determination of liability. Third, they note that individuals are rarely entirely free from fault, which makes strict liability a practical approach to address wrongdoing[564].

(2) Does Strict Liability Incentivise Harm Mitigation Initiatives?

It has been widely argued that strict liability in civil law creates a stronger incentive for manufacturers to make safer products[565]. Particularly in situations where owners and operators are unable to exercise control over an AI system, fault-based liability fails to achieve its primary objective of encouraging more cautious behaviour. This lack of control has led to the adoption of liability frameworks focused on inherent danger or strict liability, which emphasise accountability regardless of fault[566]. Therefore, it is stated that strict liability can be effectively applied in areas where the risks posed by AI-driven autonomous systems cannot be fully assessed[567]. Furthermore, it negates the necessity for legislators or courts to identify the optimal level in the design and testing of these AI systems to ascertain negligence[568].

In scholarly discourse, particularly from an economic and social welfare perspective, it has been argued that implementing strict liability instead of

562 COOPER, et al., Accountability, 2022, p. 873.
Cooper et al. do not advocate for the implementation of strict liability in criminal law but rather highlight the challenges associated with fault-based liability.
563 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 180.
564 For the assessment, see: MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 105.
565 ABBOTT, The Reasonable Computer, 2018, p. 22; BAK, Medeni Hukuk, 2018, p. 221; PAGALLO, The Laws of Robots, 2013, p. 116.
566 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 9.
567 HILGENDORF, Digitalisierung, Virtualisierung und das Recht, 2020, pp. 413-414.
568 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 13.

a fault-based regime may be more effective. Such a framework not only encourages users to adopt advanced technological solutions, but also fosters investment by firms, allowing manufacturers to see tangible benefits from their R&D expenditures. This approach is particularly relevant for fully autonomous systems, resting on the premise that victims' precautions are generally of little significance in avoiding accidents in such scenarios[569]. Indeed, under a fault-based liability regime, companies are required to compensate for damages only when their risk-taking exceeds what is considered acceptable, often involving complex calculations of risk levels. Moreover, as a rule, they can avoid liability by proving that they exercised the required standard of care. In contrast, strict liability obliges firms to compensate for all damages regardless of the level of risk, thereby simplifying the process. The application of strict liability is particularly advantageous in areas where harm occurs rarely[570]. Therefore, determining in which areas AI-driven autonomous systems are utilised, harm occurs frequently and in which areas it occurs rarely (and perhaps severely) will guide the economic-legal practice on this matter.

The prospect of being held liable for every type of harm that occurs may discourage manufacturers from taking risks, potentially hindering innovation. Such a deterrent effect could slow technological advancements and limit the development of new, potentially beneficial products and systems[571]. By contrast, an alternative viewpoint posits that imposing liability does not inherently impede innovation; rather, it can motivate companies to develop technologies that mitigate risks while enhancing the safety and reliability of their products. This strategy not only minimises the probability of harm but also fosters greater user confidence and broader acceptance of such technologies[572].

Conversely, if preventing harm from AI systems' operation requires all involved actors (such as the manufacturer, owner and operator) to exercise

569 DE CHIARA, et al., Car Accidents, 2021, p. 3, 8, 10.

570 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 13. See also: European Parliament, Artificial Intelligence and Civil Liability: A European Perspective, Policy Department for Justice, Civil Liberties and Institutional Affairs, Committee on Legal Affairs (JURI), PE 776.426, 24.07.2025, https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)776426, p. 43 ff., 68, 90 f., (accessed on 01.08.2025).

571 NISSENBAUM, Accountability in a Computerized Society, 1996, p. 39; BALKIN, The Path, 2015, p. 52; LOHMANN, Liability Issues, 2016, p. 338 f.; OSMANI, The Complexity of Criminal Liability, 2020, p. 75.

572 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 9.

due care, and the courts cannot determine the specific cause of the harm; placing sole liability on the manufacturer through strict liability may cause other actors to behave less cautiously[573]. This situation aligns with the "double moral hazard problem" described in economics literature. In a strict liability regime, where the injurer bears the entirety of the harm, the victim (operator in this case)[574] has little to no incentive to take measures to prevent the harmful outcome[575]. Therefore, the adoption of strict liability is justifiable only in areas where operators lack control over the system, making their exercise of due care ineffective.

(3) Defining the Scope of the Strict Liability Regime

The adoption of a strict liability regime for AI-driven autonomous systems provides significant legal simplicity[576]. However, given that fault-based liability is the general rule, and strict liability is the exception, it is crucial to regulate the conditions and boundaries of strict liability for hazardous activities in a manner that ensures clarity and precision, reflecting its exceptional nature. Thus, the activities falling within the scope of strict liability can be clearly identified[577].

The regulation of liability for hazardous activities, a form of strict liability, typically follows two main approaches. The first involves the enactment of specific legislation, as seen in Switzerland, to address specific sources of risk; such as motor vehicles or damages arising from the operation in nuclear facilities[578]. The second approach is the inclusion of a general provision within the civil code for strict liability, leaving the resolution of specific cases to judicial discretion based on the circumstances of each case. Furthermore, the emergence of new risk phenomena, such as those associ-

573 *Ibid*, p. 13.
574 For the purpose of this study, see, for the interpretation of the term 'operator': Chapter I, Section D: "Addressing Liability: Key Actors and Entities".
575 *Ibid*, p. 9; DI/CHEN/TALLEY, Liability Design, 2020, p. 3.
576 JANAL, Die deliktische Haftung, 2016, p. 155.
577 *Ibid*, p. 157; AKKAYAN YILDIRIM, 6098 Sayılı Türk Borçlar Kanunu, 2012, p. 211.
578 See *e.g.*: Art 3(1) of Kernenergiehaftpflichtgesetz (Swiss Federal Nuclear Energy Liability Act, KHG), enacted on 13.06.2008, in force as of 01.01.2023, last amended on 01.01.2022, https://www.fedlex.admin.ch/eli/cc/2022/43/de. (accessed on 01.08.2025).

ated with AI, may necessitate the introduction of specialised legislation to govern strict liability in these contexts[579].

As a form of manufacturer's strict liability, a specialised regime for robot product liability would be analogous to the liability framework for genetically engineered products under Section 37(2) of the German Genetic Engineering Act (GenTG)[580], which also encompasses development risks. This approach does not focus on fault-based breaches of duty but instead imposes strict liability for the utilisation of a specific technology[581]. In addition, Section 7 of the German Road Traffic Act (StVG) establishes the strict liability of the vehicle operator.

Strict liability for hazardous activities addresses the inherent risks associated with a specific activity or product. It is not feasible to assume that every manufacturing activity or product inherently entails such typical risks. However, if a product or manufacturing activity involves inherent dangers, the legislator may regulate it under the framework of strict liability for hazardous activities. One view posits that in the absence of a specific strict liability regime for AI-driven autonomous systems, such liability cannot be applied. Nevertheless, if these systems fall within the scope of existing strict liability categories, they may still be covered[582]. Despite opposing views[583], AI does not fit within frameworks such as employer's liability or liability for animal keepers. The most reasonable approaches are strict liability for hazardous activities and producer's liability; however, it is argued that both are inadequate for addressing the advanced capabilities of AI. Consequently, it is suggested that new regulatory frameworks are required[584].

Strict liability in civil law aims to strike a balance between society's need for technological innovation and the protection of individuals from harm. It ensures that the responsibility for damages caused by AI-driven systems falls not on random victims, but rather on those who economically benefit from such innovations[585]. Indeed, the essence of hazard-based liability lies

579 *Ibid*, p. 204 f.

580 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).

581 ZECH, Zivilrechtliche Haftung, 2016, p. 200.

582 BAK, Medeni Hukuk, 2018, p. 222.

583 An opinion suggests that as there is no explicit requirement for the term "employee" to refer solely to human and it could be interpreted broadly to encompass AI-driven systems. See: SELANIK, Adam Çalıştıran, 2022, p. 358.

584 BAK, Medeni Hukuk, 2018, p. 223.

585 FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 880 f.

in the inherent risks and probability of accidents: those who benefit from a hazardous activity must also bear the resulting disadvantages[586]. In this context, employing the "deep pocket" theory can be advantageous. This theory suggests that individuals or entities engaged in hazardous, yet profitable and socially beneficial activities should allocate a portion of their earnings to compensate society for any resulting damages[587]. For instance, in the case of autonomous vehicles, it has been proposed that strict liability should apply to manufacturers, either instead of or in addition to the owners, as both parties derive economic benefit from these systems[588].

Although a balance between benefit and burden is necessary, holding parties liable for all accidents somehow related to the source of danger would undermine economic viability. Therefore, liability is limited to operational risks, meaning it applies only when the damage is caused by a risk inherent to the danger[589]. This conceptualisation of danger encompasses situations with an expected potential to cause harm and outcomes directly related to the operation, whether in terms of quality or quantity. For instance, the risk of a self-driving vehicle failing to recognise a pedestrian crossing and causing bodily injury can be considered an operational risk in this field. However, an entirely unforeseeable event, such as the vehicle's software hacking into an unrelated information system, would not be considered a risk connected to the operation and therefore should not result in strict liability.

In my opinion, it is not feasible to categorise all AI-driven autonomous systems as inherently hazardous activities. Firstly, there is significant diversity among AI-driven systems, and their classification varies not only based on a risk-based approach but also according to the sectors in which they are utilised. Furthermore, the fundamental issue with AI is not its frequent or large-scale potential to cause harm, but rather the challenges that arise from its autonomy. These include reduced human control, unpredictability, and the difficulty of providing retrospective explanations. From this perspective, AI can be more accurately likened to viruses or bacteria[590] in terms of risk, rather than to a power station.

586 CHRISTALLER et al., Robotik, 2001, p. 154.
587 ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 387; OSMANI, The Complexity of Criminal Liability, 2020, pp. 68-70.
588 SEDLMAIER/KRZIC BOGATAJ, Die Haftung, 2022, p. 2955.
589 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 445.
590 See: Chapter 1, Section E(1)(f): "Lack of Predictability in AI-Driven Autonomous Systems".

(4) The EU AI Liability Directive (AILD) and Strict Liability Regime within the EU

The European Union's draft AI Liability Directive (AILD)[591], initially proposed in September 2022, sought to address the issue of non-contractual civil liability for damages caused by AI systems. The Directive was intended to complement the EU's broader AI regulatory framework, which includes the AI Regulation (AI Act) and the revised Product Liability Directive (PLD). Nevertheless, the proposal has encountered obstacles and delays: its necessity has been contested due to overlapping, particularly in light of the inclusion of software within the scope of the revised PLD, which was published in the Official Journal of the EU on 18 November 2024. In this regard, the European Commission announced in its 2025 Work Programme, published in February 2025, that it intended to withdraw the proposed AILD, citing the absence of any foreseeable agreement among institutions and stakeholders. The Commission further indicated that an alternative proposal or a different regulatory approach should be considered[592].

It is self-evident that this Directive, along with the preceding initiatives, did not extend to matters of criminal liability. Nevertheless, as will be discussed in the section addressing the EU AI Regulation (AI Act), there are certain guiding aspects pertaining to criminal liability[593]. The introduction of strict liability for AI-caused civil liability was initially proposed by the European Parliament's resolution in 2020[594]. This resolution proposed a Regulation that would take precedence over national liability regimes on the matter. Specifically, it proposed the establishment of strict liability for operators of high-risk AI systems. This would hold operators liable for harms caused by the AI's both physical and virtual activities, regardless of

591 European Commission, Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive), COM(2022) 496 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496 (accessed on 01.08.2025).

592 European Commission, Annex to the Commission Work Programme 2025, COM(2025) 45 final, 06.02.2025, https://commission.europa.eu/document/download/7617998c-86e6-4a74-b33c-249e8a7938cd_en, p. 26.

593 See: Chapter 4, Section C(5)(c)(5): "The EU AI Regulation (AI Act) and the Imposed Duty of Care".

594 European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), P9_TA(2020)0276, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html. (accessed on 01.08.2025).

whether they acted with due diligence or whether the damage resulted from autonomous AI processes. Therefore, the operators could only be exempt from liability in cases of *force majeure*. Furthermore, the resolution also envisioned empowering the Commission to maintain an exhaustive list of high-risk AI systems and critical sectors in an annex, which could be amended to include or exclude AI types or sectors based on evolving risk assessments. Additionally, it required operators to secure adequate liability insurance to cover compensation obligations[595].

Following subsequent discussions, the Commission, in the first draft of the AILD, rejected the European Parliament's view that a strict liability regime would be more appropriate. As noted in the Explanatory Memorandum, "[s]trict liability was considered disproportionate by the majority of business respondents", leading to the exclusion of strict liability for operators from the draft. Instead, it focused on facilitating fault-based liability claims by introducing measures such as a rebuttable presumption of causality and provisions for the disclosure of evidence related to high-risk AI systems.

Keeping in mind the *ex post* issues arising from the legal challenges posed by AI-driven autonomous systems[596], Article 4 of the draft AILD introduced a rebuttable presumption of causality in fault-based liability claims; applicable under specific conditions. For high-risk AI systems, the presumption applies if the claimant proves the defendant's fault (*e.g.*, non-compliance with duties under the EU AI Regulation such as inadequate data quality, transparency, oversight, or cybersecurity) and established a reasonable likelihood that the fault influenced the system's output or failure, which eventually caused the harm. However, the presumption would not apply if the defendant (operator) demonstrated that sufficient evidence is reasonably accessible to the claimant. In other (non-high risk) AI systems, the presumption would apply only if proving causality is excessively difficult. The presumption could also be rebutted by the defendant under all circumstances[597].

595 Article 4 of the Proposal for the Regulation of the European Parliament and of the Council on liability for the operation of artificial intelligence-systems, within the aforementioned Resolution.

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

597 Article 4 of the Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence.

A study on the AI Liability Directive, published by the European Parliamentary Research Service in September 2024[598] (before the withdrawal) examined the potential incorporation of a strict liability framework. As explained, this framework had been the subject of ongoing debate, particularly regarding AI systems that, when properly designed and deployed, should not cause harm. Proponents argue that strict liability promotes the optimal deployment of AI technologies, simplifies victim compensation, and ensures that those who derive economic benefits from AI systems also bear the associated risks. In contrast, critics highlight potential drawbacks; including the deterrence of AI investment within the EU, restricted access to beneficial AI technologies in critical sectors such as healthcare and education, diminished enjoyment of fundamental rights, increased frivolous litigation over non-material harms, and an undue burden on small and medium-sized enterprises (SMEs), which are central to the European AI ecosystem[599].

A subsequent study, published on July 2025 at the request of the European Parliament's Committee on Legal Affairs, provided a detailed examination of how the civil liability regime within the EU should be shaped following the withdrawal of the AILD proposal. The study concludes that the revised PLD is mainly inadequate, and reiterates concerns that the AILD would have exacerbated fragmentation by operating across 27 divergent tort law systems in the member states. It further observes that the rebuttable presumptions envisaged under the AILD would apply only if claimants satisfied heavy preconditions; thereby significantly limiting their practical utility. The study criticises the broad and insufficiently defined scopes, and warns that its reliance on shifting concepts (such as interpreting fault as a breach of AI-specific duties) would generate doctrinal confusion. In the face of such ambiguity, national courts would likely revert to existing (national) strict liability rules, making the directive largely ineffective. In light of these shortcomings, the study explicitly recommends transforming the AILD into -or replacing it with- a strict liability regime applicable to

598 European Parliamentary Research Service, Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence: Complementary impact assessment, 2024, https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf. (accessed on 01.08.2025).

599 European Parliamentary Research Service, Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence: Complementary impact assessment, 2024, https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf, p. III. (accessed on 01.08.2025).

high-risk AI systems. This approach, mirroring the 2020 draft proposed by the Parliament, would align with the AI Act's categorisation of high-risk systems and allocate all liability to a single, insurable operator (or provider and/or deployer). Only such a framework, it argues, could meaningfully achieve genuine harmonisation, ensure adequate victim compensation, and provide the legal certainty necessary to foster innovation[600].

The debate surrounding civil law strict liability for AI-caused harm, as can be seen, encompasses a complex array of economic and legal aspects. Of equal importance is the question of the global trajectory on this matter. For example, the recent AI Safety Bill (SB 1047) in California[601], which proposed a (limited) strict liability framework, highlighted the potential implications of such measures and its potential effects in the EU[602]. However, the governor's veto of the bill has raised concerns about addressing AI risks within an appropriate legal framework[603]. Legislative efforts in California are particularly crucial, given that it is home to many of the world's leading technology companies, whose practices could significantly influence the global approach to AI risks.

(5) Compatibility of Strict Liability with Criminal Law Principles

Strict liability highlights the fundamental distinction between civil law and criminal law. To address the challenges of fault-based liability in offences involving AI-driven autonomous systems, it has been proposed to adapt strict liability in criminal law to fill liability gaps and ensure accountability for harm that might otherwise be dismissed as "bad luck". Proponents argue that strict liability incentivizes greater caution and higher standards

600 European Parliament, Artificial Intelligence and Civil Liability: A European Perspective, Policy Department for Justice, Civil Liberties and Institutional Affairs, Committee on Legal Affairs (JURI), PE 776.426, 24.07.2025, https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)776426, *passim*, (accessed on 01.08.2025).

601 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, Senate Bill No:47 (SB-1047), 09.03.2024, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047. (accessed on 01.08.2025).

602 AI Liability Directive: Study of the European Parliament on AI liability, 20.09.2024, https://www.noerr.com/en/insights/ai-liability-directive-study-of-the-eu-parliament-on-ai-liability. (accessed on 01.08.2025).

603 SAMUEL Sigal/PUPER Kelsey/MATTHEWS Dylan, "California's governor has vetoed a historic AI safety bill", 29.09.2024, https://www.vox.com/future-perfect/369628/ai-safety-bill-sb-1047-gavin-newsom-california. (accessed on 01.08.2025).

of conduct, promotes procedural efficiency by simplifying liability determination, and is practical since individuals are rarely entirely free from fault[604].

The concept of strict liability in the context of criminal law is not unfamiliar within the Anglo-American legal tradition. However, it continues to be a highly contentious issue[605]. Nevertheless, this approach is largely flawed within the framework of the Continental European legal tradition, where culpability remains a cornerstone of criminal liability[606]. The adoption of strict liability principles by criminal courts in medical liability cases, originally developed in civil courts, has already been the subject of intense criticism[607]. To fill liability gaps, a criminal strict liability framework akin to that in civil law may seem effective. However, the principle of culpability remains a substantial obstacle to its adoption[608]; and in addition to existing criminal law mechanisms, this gap can be partially addressed by introducing a new endangerment offence[609].

It can be argued that negligence already serves filling the gap between intentional crimes and strict liability[610]. Furthermore, the aforementioned argument that strict liability incentivises manufacturers to reduce risks is applicable solely within the scope of civil law and does not necessitate the establishment of strict liability in criminal law. It is not necessary for them to be held strictly liable separately under both criminal and civil law. The potential for manufacturers to be held financially accountable already serves as a sufficient incentive for them to develop safer products. Undoubtedly, under criminal law, an individual can only be held responsible if fault is present, and not every act requires criminal liability. However, as

604 ZHAO, Principle of Criminal Imputation, 2024, p. 26 f.; MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 105.
See also: MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 180; COOPER, et al., Accountability, 2022, p. 873. Cooper et al. do not advocate for the implementation of strict liability in criminal law but rather highlight the challenges associated with fault-based liability.

605 CALO, Robotics and the Lessons, 2015, p. 554; GÜNSBERG, Automated Vehicles, 2022, p. 446; BALKIN, The Path, 2015, p. 52.

606 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 174.

607 WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 148.

608 DUTTGE, StGB § 15 MüKo, 2024, Rn. 105; IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 430.

609 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 451.

610 MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 106.

previously discussed[611], fault-based liability tends to create a retribution gap rather than merely a criminal liability gap. Concepts such as permissible risk and the principle of reliance, as explored later in this study, not only fail to address this gap but also indicate that certain areas might remain entirely beyond the reach of criminal liability. Therefore, solutions must be developed to address society's retributive needs adequately; otherwise, they will be disregarded altogether.

d. Product Liability

(1) Introducing Product Liability for AI-Driven Autonomous Systems

The challenges posed by AI-driven autonomous systems in terms of predictability and controllability are particularly evident in "self-learning" adaptive systems. Illustrating this issue is a case of a 14-year-old who became increasingly withdrawn and ultimately committed suicide after forming a deep emotional attachment with a character they had created on "*Character.ai*"[612] (a platform designed to build and interact with AI-generated and driven characters, allowing users to simulate conversations or storytelling experiences with personalised virtual personas)[613]. Although it must be acknowledged that, for this incident, numerous factors contributed to the process leading to the child's suicide, which makes the determination of causation and negligence challenging from a legal perspective, it is evident that similar cases involving LLM chatbots are becoming widespread. Indeed, the most frequently discussed example of this in legal literature is the *Microsoft Tay* incident[614].

In the *Character.ai* incident, the developers who created and made the platform available to the public should have implemented a range of fine-tuning measures and guardrails to prevent chatbots from generating certain types of expressions, encouraging specific harmful behaviours, and being manipulated, particularly in light of incidents such as that of

611 See: Chapter 3, Section C: "Various Liability Models for the Person Behind the Machine".
612 https://character.ai. (accessed on 01.08.2025).
613 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).
614 See: Chapter 2, Section C: "Prominent Cases Highlighting AI-Related Liability".

Microsoft Tay[615]. These measures include the curation of training data and the implementation of toxicity filters, among others. However, predicting and preventing all undesirable outputs through guardrails, especially given the existence of adversarial techniques such as prompt injection, remains unachievable.

Moreover, in such cases, where chatbots can be customised by users; the developers' responsibility to monitor the product after its release becomes significantly more challenging. In any case, manufacturers are obligated to take precautions against foreseeable and avoidable outcomes. Among other precautions, a warning on *Character.ai*, issued prior to this incident, explicitly stated that the characters' statements were entirely fictional. Following the incident, it was updated to: "*This is an AI chatbot and not a real person. Treat everything it says as fiction. What is said should not be relied upon as fact or advice*". However, such warnings may not suffice to absolve manufacturers of liability, as will be discussed below. Besides, provisions in user agreements prohibiting certain content or imposing age restrictions are neither particularly effective nor sufficient from the perspective of criminal law. At best, such provisions could be regarded as an assumption of risk or consent by the user. Even so, these principles have their boundaries.

Regarding AI-driven systems, due to the challenges in fault-based liability, the notion that society must tolerate such exceptional outcomes (as in the example of the 14-year-old child, even if the causal nexus had been clear) can be questioned; particularly as such adaptive self-learning systems become more widespread[616]. On the other hand, the application of product liability rules and the imposition of strict liability on manufacturers could be considered. In case that the definition of 'product' includes 'software'; subjecting manufacturers, who derive significant profits from these systems to at least civil liability appears justifiable from the perspective of legal policy[617].

615 See: Chapter 4, Section C(4)(a)(2): "Learning from Mistakes and Hindsight Bias". See also: HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 552-553.

616 See: Chapter 4, Section C(5): "The Permissible Risk Doctrine".

617 HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 553

(2) Responsibility Shifting to Manufacturers

In the functioning of AI-driven autonomous systems, users' control over such systems tends to diminish significantly. It would not be incorrect to assert that the degree of autonomy of these systems is inversely proportional to the level of control exercised by users (or in the case of semi-autonomous vehicles, by drivers). Consequently, in cases where a legal interest is violated involving such systems, the user's liability is limited to the extent of their control. However, the adaptability and autonomy of these systems primarily manifest during their development and design phases. For example, the ability of a semi-autonomous vehicle to accurately identify and distinguish bicycles and motorcycles in traffic is determined during the stage of training and development of their software, well before the vehicle is manufactured. Accordingly, the literature commonly observes a shift in both civil and criminal liability from users to manufacturers[618]. In this regard, traffic accidents involving such systems could potentially become a matter of product liability, where the focus shifts from misconduct to product defects[619].

Contrary to the widespread opinion, a cautious approach should be taken toward viewing occupants of self-driving vehicles as mere passengers exempt from liability. Activating such vehicles creates inherent risks and constitutes task delegation to AI-driven autonomous systems. Unless entirely passive, this activation point should be central to liability analysis[620]. As task delegation to AI systems increases, evaluating whether such delegation falls within permissible risk becomes crucial[621].

618 HILGENDORF, Teilautonome Fahrzeuge, 2015, p. 25; HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 24; SCHUSTER, Künstliche Intelligenz, 2020, p. 396; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 1122; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 286, 289; HOHENLEITNER, Die strafrechtliche Verantwortung, 2024, p. 26; REVOLIDIS/DAHI, The Peculiar Case, 2018, p. 75; SANDHERR, Strafrechtliche Fragen, 2019, p. 2 f.; HILGENDORF, Wer haftet für Roboter? Autonome Autos, in: Legal Tribune Online (LTO), 21.07.2014, https://www.lto.de/recht/hintergruende/h/autonome-autos-google-car-haftung-verkehrsrecht/. (accessed on 01.08.2025).

619 GOMILLE, Herstellerhaftung, 2016, p. 82; LOHMANN, Liability Issues, 2016, p. 337.

620 For a detailed discussion see: Chapter 4, Section C(5)(b)(3)(d): "Delegating Tasks to AI-Driven Autonomous Systems: An Alternative Approach for Liability".

621 See: Chapter 4, Section C(5): "The Permissible Risk Doctrine".

(3) The Essence of Product Liability

The earliest examples of product liability can be traced back to the *Code of Hammurabi*: "*229: If a builder build a house for some one, and does not construct it properly, and the house which he built fall in and kill its owner, then that builder shall be put to death.*"[622].

In modern manufacturing processes, numerous parties are involved in the journey of a product until it reaches the consumer. In this process, while the consumers are within a contractual relationship, third parties which cannot be addressed through the contract may also suffer harm. Product liability serves to fill this gap. Thus, modern strict product liability emerged over the past century as a response to the inadequacies of contract law and negligence principles in complex, multi-layered production and distribution chains, particularly for dangerous products. Expecting the injured end-user to bear the cost of harm arising from defective or unsafe products was deemed unfair, which led to the development of strict liability principles[623]. This approach provides greater certainty by imposing upon manufacturers a duty to compensate for damage caused by the failure of their products to meet legitimate safety expectations, and identifying in advance the party who may be held liable. Thus, it encourages manufacturers to improve product safety with clarity and ultimately aims to protect the persons and property adversely affected by defective products[624].

Under German civil law, various types of liability may apply to damage caused by AI-driven autonomous systems. For example, contractual liability, statutory liability under Section 823 of the German Civil Code (*Bürgerliches Gesetzbuch*), the owner's compensation obligation under Section 7 of the Road Traffic Act (*Straßenverkehrsgesetz*), and product liability are all potentially applicable[625]. However, in production and distribution chains involving multiple parties, identifying the bases of harm caused by a product and determining that it arises from the fault of a particular party can be exceedingly difficult[626].

622 Code of Hammurabi (c. 1700 B.C.E.) Yale Law School, Translation: L. W. King, https://avalon.law.yale.edu/ancient/hamframe.asp (accessed on 01.08.2025). See: NISSENBAUM, Accountability in a Computerized Society, 1996, p. 25.
623 KARNOW, The application, 2016, pp. 65-66.
624 FUCHS/BAUMGÄRTNER, Ansprüche aus Produzentenhaftung, 2011, p. 1061; TURNER, Regulating AI, 2019, p. 94 f.
625 HILGENDORF, Robotik, Künstliche Intelligenz, Ethik und Recht, 2020, p. 551 f.
626 HAGER, Umwelthaftung, 1990, p. 398.

To address these challenges, civil product liability has developed as a form of strict liability, significantly influenced by the possibility to insure against such risks. Thus, under German law, product liability is considered as a form of strict liability that incorporates elements of fault[627]. The German Product Liability Act (*Produkthaftungsgesetz* - ProdHaftG)[628], being the primary source of product liability in German law, is an implementation of the 1985 EU Product Liability Directive (Directive 85/374/EEC)[629], which holds manufacturers strictly liable for defective products that cause injury[630]. Moreover, the provision in Section 15(2) of the ProdHaftG clarifies that the application of other types of liability is not precluded. Therefore, fault-based producer liability pursuant to Section 823(1) of the BGB, which constitutes a specific form of the general duty to ensure safety, further developed and shaped by case law to address modern industrial production, continues to apply[631].

(4) Manufacturer's Duties

It should initially be stated that product liability can arise in three distinct forms: design defects, manufacturing defects, and failure to provide adequate instructions and warnings. A design defect exists when a product, at the time it is placed on the market, falls short of the prevailing state of the art and fails to meet the required safety standards. In such cases, foreseeability may play a role; the harm could have been avoided if the product had been designed differently. However, a risk-benefit analysis is typically conducted, as it is neither practical nor economically feasible for

---

See: Chapter 4, Section D(2)(b)(1): "Liability Challenges in the Production Chain of AI-Driven Autonomous Systems".

627 ZECH, Gefährdungshaftung, 2013, p. 23; FUCHS/BAUMGÄRTNER, Ansprüche aus Produzentenhaftung, 2011, p. 1061.

628 Gesetz über die Haftung für fehlerhafte Produkte (ProdHaftG), enacted on 15.12.1989, last amended on 23.11.2022, https://www.gesetze-im-internet.de/prodhaftg/BJNR021980989.html. (accessed on 01.08.2025).

629 Council of the European Communities, Council Directive 85/374/EEC of 25 July 1985 on the Approximation of the Laws, Regulations, and Administrative Provisions of the Member States Concerning Liability for Defective Products, OJ L 210, 07.08.1985, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31985L0374. (accessed on 01.08.2025).

630 OSMANI, The Complexity of Criminal Liability, 2020, p. 62; BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 4.

631 SEUFERT, Wer fährt, 2022, p. 321.

all products to be made for example, from exceptionally durable materials, solely to prevent damage. Hence, a design defect pertains to flaws in the product's design, which inevitably affect the entire series during mass production. By contrast, a manufacturing defect arises when the product is designed without fault but deviates unintentionally from the quality standards intended by the manufacturer during the production process[632]. Furthermore, in mass production, individual outliers (*Ausreißer*) may also occur[633].

Products that are free from design or manufacturing defects and have undergone sufficient testing typically do not cause harm when used as intended. Nonetheless, the manufacturer is obligated to provide clear instructions for use and to inform consumers about the known risks of foreseeable misuse as well as unknown potential dangers[634]. With respect to product warnings, the manufacturer must also identify the target audience for the product and issue warnings that are tailored to that specific user group[635].

The manufacturer is responsible for any safety deficiencies that are known or reasonably knowable at the time the product is released on the market. However, the manufacturer's obligation of due diligence does not end upon the release of the product. For instance, they must continue to fulfil their obligations by providing security updates and actively monitoring the product to identify any previously unknown risks[636]. This obligation of due diligence imposes both passive obligations, such as receiving user complaints, and active obligations, including evaluating such data and taking appropriate action where necessary[637]. The active monitoring requirement is particularly critical for high-risk AI (-driven) systems. To meet these obligations, manufacturers may establish operational facilities dedicated to collecting and evaluating information regarding the product's real-world performance[638]. If, through such mechanisms, the manufacturer becomes aware of a product's dangers, they are obliged to take corrective

632 GOMILLE, Herstellerhaftung, 2016, p. 77, KARNOW, The application, 2016, p. 66 f.
633 FUCHS/BAUMGÄRTNER, Ansprüche aus Produzentenhaftung, 2011, p. 1059.
634 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 441; KARNOW, The application, 2016, pp. 66-67; VOGT, Fahrerassistenzsysteme, 2003, p. 159.
635 Von WESTPHALEN, Das neue Produkthaftungsgesetz, 1990, p. 88.
636 RAUE, Haftung, 2017, pp. 1843-1846.
637 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 441.
638 KULLMANN, Produkthaftung, 2002, p. 6; VOGT, Fahrerassistenzsysteme, 2003, p. 159, SANDER/HÖLLERING, Strafrechtliche Verantwortlichkeit, 2017, p. 197.

measures, which may include modifying the production process, issuing warnings to consumers, or initiating a recall if required[639].

Article 7 of the new (revised) EU Product Liability Directive of 2024 (PLD)[640] defines defectiveness and specifies the factors to be taken into consideration. Accordingly, "[a] product shall be considered defective where it does not provide the safety that a person is entitled to expect or that is required under Union or national law". The assessment of whether a product is defective occurs when it is released on the market. Additionally, according to Article 7(3), "[a] product shall not be considered to be defective for the sole reason that a better product, including updates or upgrades for a product, has already been or is subsequently placed on the market or put into service"[641]. Moreover, for certain products, additional safety-enhancing measures can be made available for purchase separately, particularly in terms of price-performance considerations[642].

According to Section 1(2)(4) of the ProdHaftG and Article 11(1)(d) of the new PLD where the defectiveness that caused the damage is due to the product's compliance of the product with legal requirements, liability is exempted. In this regard, standards play a crucial role; failure to comply with technical norms generally signifies a product defect. However, it is argued that, if the *state of science and technology* evolves beyond these standards, the most advanced state becomes applicable. In such cases, these standards represent only a minimum threshold, and additional obligations may arise to reflect the latest advancements[643]. In criminal law, particularly in the context of negligence, the duty of care and the general principle of not causing harm may require exceeding established standards. Consequently,

639 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 441.

640 European Union Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products. Official Journal of the European Union L 275, 28.10.2024, https://eur-lex.europa.eu/eli/dir/2024/2853/oj. (accessed on 01.08.2025).

641 For an evaluation, see: SCHRADER, Haftungsfragen, 2016, p. 242.

642 Von WESTPHALEN, Das neue Produkthaftungsgesetz, 1990, p. 88.

643 SEUFERT, Wer fährt, 2022, pp. 322-323.
Additionally, tort producer liability considers public product safety law as a minimum standard for determining the duty of care, meaning that compliance does not absolve manufacturers from addressing additional risks. For AI products particularly those with low or minimal risks, civil courts may need to develop specific standards based on general product safety laws, such as Section 3 of the ProdSG (Produktsicherheitsgesetz), to address gaps in existing regulations. See: IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 293 f.

this may not always lead to the same outcomes as those reached under civil law.

In the context of legitimate safety expectations, in the *Airbag* decision, the German Federal Court of Justice (BGH) held that the greater the danger posed by a product, the higher the obligations placed on the manufacturer[644]. Specifically, manufacturers are required to mitigate risks of malfunction through design measures, provided that such measures are within the bounds of technical feasibility and economic reasonableness. Therefore, for autonomous vehicles, safety expectations would be exceptionally higher due to the significant risks to life and health, as well as the increased likelihood of damage arising from their operation in complex traffic environments[645]. However, according to Section 1(2)(5) of the ProdHaftG or Article 11(1)(e) of the new PLD, if risks associated with a product cannot be avoided through the state of science and technology, or if such measures are unreasonable for the manufacturer, the product can still be marketed after weighing the remaining risks against the benefits. If this assessment concludes that the product can be marketed, the manufacturer is then obligated to provide instructions regarding the unavoidable risks inherent in the product's design. This allows the consumer / user to decide whether to use the product and whether the benefits outweigh the associated risks[646].

(5) Specific Challenges for AI-Driven Systems in Product Liability

Three main issues arise in the context of product liability for AI-driven systems. First, there is the challenge of defining AI as a 'product' within this framework. Second, the interpretation and scope of 'defect' in AI-driven autonomous systems requires careful analysis, since traditional definitions may not encompass the unique, evolving characteristics of such systems, as

644 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.

645 SCHRADER, Haftungsfragen, 2016, p. 243.

646 VOGEL/BÜLTE, § 15 Vorsätzliches fahrlässiges Handeln in LK, 2020, p. 1186, Rn. 279; FUCHS/BAUMGÄRTNER, Ansprüche aus Produzentenhaftung, 2011, p. 1058.
See also: HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 441.
For the view that liability exemption in favour of particularly autonomous vehicle manufacturers should be out of the question because it would undermine their incentives to produce error-free products; see: WAGNER, Produkthaftung für autonome Systeme, 2017, p. 762.

exemplified above. Finally, the burden of proof poses significant challenges, particularly given the inherent opacity of many AI systems, often described as the 'black box' problem[647].

Firstly, under the current legal framework of the German ProdHaftG (Section 2) and the EU Product Liability Directive (PLD) of 1985 (Art. 2), a 'product' is defined as any movable item, even if it forms part of another movable or immovable item. As a result, software does not fall within this definition. This issue has been subject to extensive debate in legal literature. Considering their earlier date, the original rationale of limiting the definition of 'product' in these provisions were to exclude buildings and land from their scope[648].

Nevertheless, software stored on a physical data carrier, or integrated into a final product where it functions as a tailored component and where the manufacturer is responsible for its installation and updates, may be deemed tangible; and thus, fall within the scope of product liability[649]. Moreover, AI systems can also be offered as a service[650]. However, the situation is less clear when software is downloaded independently or is not embodied but is stored in the cloud and accessible only via the internet, especially considering that electricity is explicitly specified as an exception[651]. Consequently, if harm is caused by an embodied robot due to a recent separate software update, civil product liability would not have applied under the previous legal regime[652]. In cases where AI (systems) are not classified as product, manufacturers of autonomous vehicles, for instance, could limit their liability under the product liability law by exclusively offering potentially problematic software (prone to errors) through user requested updates rather than integrating it into the product at the time of sale[653]. Nevertheless, this will no longer pose an issue, as Article 4(1) of the new EU PLD of 2024 has expanded the definition of product to include software[654].

---

647 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 13.
648 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 443.
649 HILGENDORF, Digitalisierung, Virtualisierung und das Recht, 2020, p. 414; CHANNON/MARSON, The Liability for Cybersecurity, 2021, p. 5.
650 BUITEN/DE STREEL/PEITZ, The Law and Economics of AI Liability, 2023, p. 5.
651 SEDLMAIER/KRZIC BOGATAJ, Die Haftung, 2022, p. 2955.
652 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 443; SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 260.
653 VELLINGA, Cyber Security, 2023, p. 134.
654 Pursuant to Recital 13, AI systems made available through a software-as-a-service (SaaS) model also qualify as product. However, it appears that debates will persist,

The second issue concerning AI-driven systems relates to the determination of 'defect'. Indeed, the product liability model may be well-suited for simple automated systems; but AI-driven autonomous systems may generate unforeseeable outcomes and involve unrecognisable dangers due to their inherent *ex ante* uncertainties[655]. While their adaptive nature is a desirable feature, this same characteristic may lead to violations of legal interests. Therefore, defect or malfunction cannot be understood in the conventional sense[656]. Failures of these systems typically arise from a combination of limitations in the learning process rather than from inherent defects. These systems may fully comply with legal requirements but still fail to function within the parameters set by their design and training[657]. Furthermore, another significant challenge in applying product liability arises from the difficulty in determining whether a product became defective due to its self-learning and adaptive capabilities after leaving the control of the manufacturer or developer, as well as whether the issue originated from these features[658]. In this regard, Article 7(2)(c) of the revised PLD explicitly states that "the effect on the product of any ability to continue to learn or acquire new features after it is placed on the market or put into service" shall also be taken into account in assessing the defectiveness of the product.

The third issue regarding such systems is the burden of proof. The EU sets general rules for high-risk AI systems and lets relevant standardisation organisations establish detailed standards. Thus, the new EU legislation aims to facilitate the process for individuals to hold AI developers liable in instances of AI "malfunction" in civil product liability cases. These regulations will apply both in situations of fault-based liability and liability for defects under product liability. By adjusting the rules of evidence, the EU intends to simplify the process for injured parties to substantiate their

particularly regarding whether certain types of software updates and upgrades should be classified as services or as products.

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

656 MILLAR/KERR, Delegation, 2016, p. 124.

657 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).

658 OSMANI, The Complexity of Criminal Liability, 2020, p. 56; ČERKA/GRIGIENĖ/SIRBIKYTĖ, Liability for Damages, 2015, p. 386.

claims, thereby imposing greater accountability on AI developers[659]. However, as highlighted by the opaque nature of machine learning models[660], proving product defects in such systems would be extremely challenging[661].

According to one perspective, opacity surrounding technical products and consumer trust often serves as a basis for establishing a protective guarantor position for manufacturers. Given that the end-user has less knowledge of the system's complexities compared to the manufacturer, and that the manufacturer is better positioned to understand and anticipate the product's risks, their role as a guarantor (entailing a duty of care and a continuing obligation to monitor the product even after it enters the market) is important. This approach aligns with the constitutional right to innovate and to derive economic benefits from such innovations, ensuring that the risks generated by the innovation are adequately addressed. Since the producer is uniquely positioned to understand the risks and potential harm associated with their product, despite its inherent opacity, imposing such obligations represents a reasonable risk management policy[662].

(6) Criminal Product Liability

(a) The Rationale Behind Criminal Product Liability

After *Ulrich Beck*'s influential 1986 work, "*Risikogesellschaft: Auf dem Weg in eine andere Moderne*" (Risk Society: Towards a New Modernity), and the subsequent debates it sparked, the effectiveness of criminal law as a mechanism for addressing various risks, including those arising from product defects capable of causing harm to individuals, has been a consistent focus of scholarly debate. However, the concept of "risk criminal law" which stretches traditional legal frameworks, has been criticised for raising concerns from the perspective of the rule of law. Nonetheless, criminal

659 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 295.
660 See: Chapter 1, Section E(2): "Ex Post: Opacity and Explainability in AI Systems".
661 European Parliamentary Research Service, A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles. European Parliament, 2018, https://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf, 2018, p. 26. (accessed on 01.08.2025).
662 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, pp. 302-305.

product liability is not considered to be a direct reflection of risk criminal law[663].

While product liability under civil law has long been the subject of extensive discussion and has been explicitly codified by legislative measures, the liability of manufacturers and distributors of hazardous products has received comparatively little attention within criminal law[664]. It only began to emerge as a distinct area of discourse at a later stage. Judicial decisions addressing criminal product liability remain relatively rare, as the majority of claims for damages are typically resolved through civil law mechanisms. Consequently, there has often been no perceived necessity for pursuing criminal prosecution in addition to civil remedies[665]. Thus, criminal product liability is (still) a relatively novel concept evolving in diverse ways across different jurisdictions[666]. German law does not have a distinct legal framework specifically addressing criminal product liability, and criminal liability is established under the general provisions of criminal law[667].

In cases where harm occurs due to the defects, risks, or hazardous nature of a product; liability would not only fall under product liability within the scope of civil law but could also give rise to criminal liability[668]. Thus, for potentially dangerous products, due diligence obligations imposed on manufacturers have, through case law, been extended from the domain of civil law to that of criminal law[669]. Nevertheless, unlike civil law, criminal product liability necessitates proving fault. Additionally, legal entities or partnerships with legal status cannot be held liable under criminal law[670]. However, in the context of criminal product liability, establishing a causal nexus or identifying a breach of duty of care is often difficult, which can sometimes result in impunity[671].

Criminal product liability refers to the legal liability from engaging in risky behaviour associated with products, as well as for any harm caused,

663 For a further evaluation, see: HILGENDORF, Gibt es ein Strafrecht der Risikogesellschaft, 1993, p. 15 f.
664 *Ibid*, p. 15; TIEDEMANN, Fragen, 1990, p. 2051.
665 WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 203.
666 KUHLEN, Grundfragen, 1994, p. 1142.
667 SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1937; ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 171.
668 ASARO, A Body to Kick, 2012, p. 171.
669 GLESS/WEIGEND, Intelligente Agenten, 2014, p. 582; LIMA, Could AI, 2018, p. 693.
670 GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 567.
671 KUHLEN, Grundfragen, 1994, p. 1144.

particularly to users of the product or individuals who come into contact with it. This form of liability may be invoked either through the act of introducing a product into the market or through subsequent conduct related to the product[672]. The development of criminal product liability significantly contributes to reducing inappropriate risky behaviour and motivates manufacturers to enhance product safety. Thus, it protects legal interests, particularly the life and physical integrity of consumers and others[673].

(b) General Duties of Manufacturers in the Context of Criminal Product Liability

Negligence requires a breach of the duty of care, as well as the foreseeability and avoidability of harm. In the context of criminal product liability, the manufacturer's negligent liability primarily aligns with the duty of care expected under the framework of civil law product liability, including design, manufacturing, and instruction obligations[674]. However, it is not entirely identical, as the functions of criminal and civil law diverge. Civil law is primarily compensatory in nature, focusing on redressing harm suffered by victims, whereas criminal law seeks to punish wrongdoing and deter future misconduct[675].

Determining negligence in failing to foresee risks is inherently challenging. While such assessments are typically based on industry standards and similar benchmarks[676], this becomes particularly complex in the context of AI-driven systems[677]. To exercise due care, a manufacturer must only bring products to market that correspond to the appropriate safety measures and have undergone proper testing. Even after a product has been placed on the market, the manufacturer must actively and continuously monitor the product (for example based on feedback from consumers). When un-

672 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 219.
673 KUHLEN, Grundfragen, 1994, p. 1143; THOMMEN/MATJAZ, Die Fahrlässigkeit, 2017, p. 295.
674 SCHUSTER, Strafrechtliche Verantwortlichkeit, 2019, p. 8; KUHLEN, Grundfragen, 1994, p. 1146; SCHULZ, Verantwortlichkeit, 2015, p. 194.
675 SCHUSTER, Künstliche Intelligenz, 2020, p. 397; IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 297.
676 ASARO, A Body to Kick, 2012, p. 171.
677 See: Chapter 4, Section C(4)(a): "The Boundaries of Foreseeability".

expected damage or dangers appear, the manufacturer is obliged to warn consumers and, if necessary, to recall the product[678].

Although the legal system may tolerate certain inherent design or software flaws[679], it cannot completely exonerate manufacturers who profit from sales from their criminal liability after the product enters the market[680]. Therefore, the impact of criminal product liability becomes particularly significant after the product has been placed on the market[681]. As part of their ongoing duty to monitor and track the products, manufacturers are required to identify and address potential risks that were not previously known at the time of the product's release. If new information regarding previously unidentified risks emerges, they must take appropriate measures to protect consumers and third parties[682]. This obligation arises from the manufacturer's or seller's guarantor responsibility, which imposes a duty to prevent harm associated with offering a product that poses potential dangers. A breach of this obligation may lead to criminal liability[683].

Beyond the manufacturer's specific obligations, criminal liability is fundamentally premised on the ability to foresee and control outcomes. Therefore, a manufacturer should not be held criminally liable for damages that are unforeseeable or beyond their control, particularly those arising from the actions of the product user. Assigning criminal liability for such supervening consequences would not align with the preventive or deterrent objectives of criminal law[684]. Nevertheless, the degree of control varies from case to case, which requires careful evaluation. Manufacturers may need to anticipate certain user tendencies and even misuse when designing their products. For example, while a desk is only required to bear the weight of a person leaning against it while sitting (under normal conditions), it is not designed to function as a platform for carrying heavy objects. Nonetheless, it is common for people to place heavy items on desks or even sit on them.

678 GLESS/WEIGEND, Intelligente Agenten, 2014, p. 582; WIGGER, Automatisiertes Fahren und Strafrecht, 2020, p. 201 ff.
679 See: Chapter 4, Section C(5): "The Permissible Risk Doctrine".
680 SCHUSTER, Künstliche Intelligenz, 2020, p. 398.
681 FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 884.
682 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).
683 SCHULZ, Verantwortlichkeit, 2015, p. 195; DEMIREL, Otonom, 2024, p. 1274.
684 HILGENDORF, Wozu Brauchen Wir, 2004, p. 45 f.

In such circumstances, it would be difficult to argue that a desk incapable of withstanding these foreseeable uses does not constitute a design defect[685].

(c) Key Judicial Decisions Shaping Criminal Product Liability

Several key judicial decisions have been instrumental in delineating the scope and characteristics of criminal product liability, as well as in shaping legislative efforts aimed at risk prevention. For instance, the *Contergan* (*Thalidomide)* case[686], one of the most prominent cases in the context of criminal product liability, focused on the criminal liability of a pharmaceutical manufacturer for birth defects caused by their medication. This case led to the establishment of comprehensive drug legislation in Germany, designed to enhance pharmaceutical safety and safeguard public health[687].

Another key decision, in the *Lederspray* case[688], involved a manufacturer whose leather spray product caused severe respiratory illnesses and fatalities among consumers due to its toxic composition and inadequate warnings. Although, primarily, a civil law case, it has significantly influenced discussions on criminal product liability by emphasising the critical role of proactive and continuous risk assessment and management by manufacturers. It raised important questions not only concerning the guarantor's position and corresponding active obligations, but also regarding the scope of a manufacturer's duty of care and the criteria for defining and determining negligence in the context of product safety. In criminal law, this would translate to whether the manufacturer's failure to anticipate harm or to act upon knowledge of potential risks constitutes a breach of the duty of care sufficient to support criminal liability[689].

Furthermore, due to the distinctions between criminal and civil liability, it is necessary to establish that the harmful outcome can be attributed to an

685 Addressing this issue, Article 7(2)(b) of the new PLD provides that the reasonably foreseeable use of the product shall also be taken into account in assessing its defectiveness. For the discussion see: Chapter 4, Section D(2)(c)(2): "Should Autonomous Systems Rely on Humans?".

686 Regional Court of Aachen (LG Aachen), decision of 18.12.1970, Case No. 4 KMs 1/68, 15–115/67, (*Contergan - Thalidomide case*) reported in JZ 1971, p. 507 ff.

687 KAUFMANN, Tatbestandsmäßigkeit, 1971, p. 569 f.; ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 170.

688 Federal Court of Justice (BGH), judgment of 06.07.1990, Case No. 2 StR 549/89, (*Lederspray case*), reported in NJW 1990, p. 2562.

689 HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, pp. 448-449.

individual (rather than a corporation) and that it arose from their culpable behaviour. In this regard, as it can be observed in the *Lederspray* case, the determination of criminal product liability involves a two-step analysis (company-related duties of conduct and individual duties of care): First, the conduct of the manufacturing organisation is assessed to determine whether it was causally connected to the harm and whether it constituted a breach of any (guarantor) obligations. Second, an assessment is made to determine whether the harmful outcome is attributable to an individual, based on their role and the organisational structure of responsibilities[690].

Indeed, even in the context of the problem of many hands[691], it is not the entity itself but rather the individual within the organisation who engages in culpable behaviour that becomes the subject of criminal punishment. Nevertheless, the standards of conduct applicable to the collective are not unrelated to the individual's breach of duty. Liability can only be imposed on a person if they have violated their duty of care and their behaviour meets all the conditions necessary for the imposition of liability[692].

As outlined previously, while a product may initially meet the conditions necessary for its placement on the market, it may later appear that the product carries unrecognised dangers. When reports or suspicions arise suggesting potential threats to human health, manufacturers are obliged to take appropriate action. In this regard, the German Federal Court of Justice (BGH), in the *Lederspray* case, characterised the initial act of introducing the product to the market as an active behaviour, while treating the failure to respond adequately to subsequent health risk warnings as an omission. In this context, the BGH, following deliberations on product risks during a crisis meeting, held that the failure to issue a product recall constituted omission, and a breach. This duty arose from the manufacturer's role in bringing a dangerous product into circulation, thereby imposing on them a guarantor's responsibility; because any party that places defective products on the market and creates risks for consumers is bound by such a duty to

690 KUHLEN, Grundfragen, 1994, p. 1144; SCHMIDT-SALZER, Strafrechtliche Produktverantwortung, 1988, p. 1939.

691 For the attribution of liability to an individual for criminal offences involving multiple actors, see: Chapter 4, Section D(1): "The Concept of "the Problem of Many Hands"".

692 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 297.

take preventive measures to avert harm stemming from those products[693]. Therefore, the argument that the product has left the manufacturer's sphere of control cannot be accepted[694].

In this context, a further significant challenge arises in determining whether the harm was in fact caused by the product in question. The establishment of causation can be particularly difficult, as an outcome may be correlated with multiple factors, but correlation does not necessarily imply causation. Furthermore, multiple causal factors may concurrently contribute to the harmful outcome. This complexity makes determining liability exceedingly challenging.

In the examination of criminal liability, the retrospective analysis typically begins by determining whether the company as the manufacturer breached a duty of care, and whether this breach was itself causal for the result[695]. However, one of the most debated aspects of the *Lederspray* decision (and similar cases such as *Contergan*), concerns the difficulties in establishing whether the product itself, and the failure to recall it, were genuinely causal to the health issues reported. In its decision, the BGH faced the challenge of insufficient scientific evidence to establish specific causality. When addressing this issue, the court assessed the foreseeability of harm retrospectively, based on the conditions at the time[696]. The BGH affirmed causality through a framework of general causality, ruling that causation is deemed established (even in the absence of 100% scientific proof) when all other plausible causes of harm are excluded. According to the BGH, the mere suspicion of a serious risk was sufficient to trigger the manufacturer's duty to ensure that consumers of leather sprays are protected from any potential damage to health arising from their use[697].

693 KUHLEN, Grundfragen, 1994, p. 1144; IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 299; KASPAR/REINBACHER, Fall 1: Lederspray, 2023, p. 16 f. Rn. 8 ff.

694 ROSENAU, Strafrechtliche Produkthaftung, 2014, p. 178.
See also: HILGENDORF, Zivil- und strafrechtliche Haftung, 2019, p. 450.

695 IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 298.

696 WALTER, Vorbemerkungen zu den §§ 13 ff in LK, 2020, p. 822, Rn. 90.

697 KUHLEN, Grundfragen, 1994, p. 1146; HILGENDORF, Fragen der Kausalität, 1994, p. 561; KASPAR/REINBACHER, Fall 1: Lederspray, 2023, p. 15 f. Rn. 6 ff.; GROPP/SINN, § 4 Tatbestandsmäßigkeit in Strafrecht AT, 2020, p. 150, Rn. 43 f.; WESSELS/BEULKE/SATZGER, Strafrecht AT, 2020, Rn. 229.

(d) Unique Challenges of AI Products and Criminal Product Liability

Similar to other product liability cases, proof of causation remains one of the most significant challenges in the context of AI-driven autonomous systems; a matter thoroughly analysed in this study with a particular focus on negligent liability and illustrated through concrete examples. Undoubtedly, the risk forecast associated with AI products is likely to be lower compared to traditional technical products and the principles developed for physical products may not be sufficient[698]. However, AI products within the framework of criminal product liability does not necessitate the creation of a new product category for liability purposes. Instead, the integration of AI-driven systems into pre-existing product categories enhances the established elements of "trans-classic technology"[699].

According to one perspective, AI-driven systems exhibit clear distinctions from pharmaceuticals and chemical substances in the context of product liability. Therefore, it diverges from the cases such as *Contergan*, *Monza-Steel*, and *Lederspray*. The fundamental distinction lies in the fact that, compared to pharmaceuticals and chemical substances, AI products are more clear-cut and controllable in terms of the separation of form and context, containment, predictability, repeatability, and troubleshooting[700].

Furthermore, foreseeing the risks associated with advanced AI-driven systems that are capable of complex interactions with people and environments and potentially "learning" and evolving beyond their original programming, may prove extremely challenging[701]. When adaptive systems of this nature are developed and made available for public use, a question arises: can they be considered defective products if they fail to function properly due to erroneous "learning"? Technically speaking, such developments need not always stem from "learning" in the strict sense; rather, undesirable outcomes may also arise from interactions with users or third parties, as is the case with chatbots. For instance, a recommendation system may inadvertently suggest harmful or inappropriate content due to patterns in user behaviour, even in the absence of adaptive learning mechanisms. From a criminal law perspective, the liability of the manufacturer could be evaluated to include the erroneous learning of self-learning systems. It

698 SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 259; IBOLD, Künstliche Intelligenz und Strafrecht, 2024, p. 246.
699 *Ibid*, p. 229, 247.
700 *Ibid*, pp. 227-230.
701 ASARO, A Body to Kick, 2012, p. 171.

may be argued that manufacturers should consider restricting the learning capacity of such systems at the time of market release. If such a limitation on the learning capacity is not provided, despite its feasibility and the reasonable expectation that the manufacturer should have implemented it, this could indicate negligence on the part of the manufacturer[702]. Nonetheless, a generalised approach in this regard would be inappropriate; instead, assessments should be made with reference to the specific system, taking into account its contextual use and intended functionality. This is because imposing temporal limitations on a system's "learning" and adaptive capacities may, to some extent, compromise the very functionalities that such technologies are designed to deliver.

To prevent harmful outcomes of this nature, programmers and maintenance personnel are held to a higher standard of care under criminal law due to their specialised technical expertise. Their unique capability to evaluate and mitigate the potential dangers associated with AI-driven systems places an increased responsibility on them. Moreover, such errors can usually be corrected by an update that can be quickly made available to all users, thereby reducing the risks and demonstrating the due diligence of those responsible for maintaining and monitoring the system[703]. Nevertheless, particularly in emerging technologies such as AI, identifying negligent conduct in risk assessment is inherently challenging. While such determinations are often made with reference to the prevailing industry standards and similar benchmarks[704], compliance with these does not necessarily equate to the fulfilment of the duty of care required under criminal negligence liability in all circumstances.

Finally, irrespective of civil product liability, the distinction between harm caused by an embodied object or by software is irrelevant in the context of fault-based liabilities. For example, in the case of offences such as negligent homicide or bodily harm under Section 222 and 229 of the StGB, this distinction holds no significance. Consequently, a manufacturer's criminal liability must apply to AI products, regardless of whether they are embodied or purely software based[705]. Manufacturers bear a critical responsibility to market products only that have undergone rigorous safety testing and adhere to the prevailing state of the science and technology[706].

702 HILGENDORF, Verantwortung im Straßenverkehr, 2019, p. 155 f.
703 SCHMIDT/SCHÄFER, Es ist schuld?, 2021, p. 418; RAUE, Haftung, 2017, p. 1843.
704 ASARO, A Body to Kick, 2012, p. 171.
705 SCHÄFER, Artificial Intelligence und Strafrecht, 2024, pp. 261-262.
706 GLESS/JANAL, Hochautomatisiertes und autonomes Autofahren, 2016, p. 565.

Furthermore, they must actively fulfil all monitoring obligations and ensure that they thoroughly fulfil their duty to instruct. This includes providing comprehensive information on both known and unknown potential risks associated with the product. In cases where a hazard is suspected, the manufacturer must address the issue promptly, maintain the safety of the product through updates and, if necessary, issue a recall. Such a proactive approach is essential to ensure the continued safety and reliability of AI products.

2. Indirect Perpetration

a. Pro Arguments for Indirect Perpetration in AI-Driven Autonomous Systems

Scholarly discourse in criminal law has seen a significant number of scholars argue that the doctrine of indirect perpetration may be applicable in cases involving the commission of criminal offences through AI-driven autonomous systems. Upon examining the origins of these perspectives, it becomes evident that they were first articulated in *Hallevy*'s works[707]. These views, which propose recognising AI-driven systems as "innocent agents" have been advocated not only within Anglo-American legal frameworks but also in Turkish and German legal systems. In this regard, the applicability of this model shall be examined.

According to *Hallevy*, AI entities are regarded as innocent agents, and their "actions" can be attributed to the individual controlling them under the "perpetration by another" model. This is analogous to the actions of a person with mental illness, where the absence of the necessary traits for criminal responsibility exonerates the perpetrator, transferring liability to the person in the background. He emphasises that this liability model does not ascribe any mental capacity (let alone human-like mental capacity) to the AI entity. He equates the use of an AI system to using an animal or a tool, such as a screwdriver, as an instrument for committing a crime, and argues that a screwdriver's "action" is, in essence, the action of the person wielding it. According to him, if the AI entity in question was more

707 The earliest source I have been able to identify is Hallevy's 2010 study; however, it is possible that earlier works on the subject may also exist: HALLEVY, The Criminal Liability, 2010.

complex -such that it decided to commit an offence based on its own accumulated experience or knowledge- or if the AI was not an innocent agent but rather a semi-innocent agent; the perpetration by another model would no longer be applicable. An example provided for this is an AI-driven autonomous robot programmed to set a factory on fire at night when no one is present or to follow its owner's commands by attacking individuals attempting to break into the owner's home[708].

According to the perspective presented here, AI can qualify as an innocent agent, and innocent agents do not necessarily have to be mere tools. An entity with some level of intelligence, such as a child, can also be regarded as an innocent agent. This "will-less tool" acting as an intermediary, does not commit the act with intent and does not need to possess culpability[709]. This is similar to using a child to pour a drug into someone else's drink[710].

Among its proponents, there is no consensus on whether this model could be applied to high-level or low-level autonomous systems. Some contend that this model is suitable only for low-level autonomous systems and is inapplicable to highly or fully autonomous systems[711]. If truly autonomous and intelligent robots were to exist, for instance, a military officer operating advanced systems such as combat drones could not be considered as an indirect perpetrator. This is because, in such a scenario, the drone, functioning as a culpable agent with control over the act, could be incriminated; although the law does not entirely preclude the application of the indirect perpetration model, particularly where the intermediary's error has been exploited[712]. Further views suggest that the perpetration by another model can only be applied if the AI is completely dependent on the person behind the machine[713] and functions solely as an instrument, lacking any capacity for self-determination[714].

---

708 HALLEVY, The Criminal Liability, 2010, p. 180 f.; HALLEVY, Liability for Crimes Involving AI, 2015, p. 41

709 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 179.

710 TURNER, Regulating AI, 2019, pp. 118 - 119.

711 HALLEVY, The Criminal Liability, 2010, p. 181

712 JOERDEN, Zur strafrechtlichen, 2020, p. 303.

713 The use of AI-driven systems as a tool is not limited to programmers; it encompasses all individuals who possess the capability and intent to manipulate and control such systems. For instance, individuals who manipulate a self-driving vehicle by providing false external inputs to trick it into accelerating could also fall within this scope. MÜSLÜM, Artificial Intelligence, 2023, p. 137.

714 FREITAS/ANDRADE/NOVAIS, Criminal Liability of Autonomous Agents, 2014, p. 150.

Some other opinions, on the other hand, suggest a distinction between the use of fully autonomous and semi-autonomous systems in the commission of crimes. When non-fully autonomous (weak AI) systems are employed, the AI system implements the intentions of the person behind it -not because it has been deceived or fails to comprehend the nature of its conduct (as a weak AI, it cannot)- but because it has been directly programmed or prompted to commit the crime. In this context, the AI system is nothing more than a more advanced, yet lifeless tool compared to traditional computers. In the case of using entirely autonomous systems on the other hand, the application of the indirect perpetration model may come into question, but only if their own criminal liability has not been recognised[715]. A similar argument suggests that for the perpetration by another liability model to be relevant in the context of AI, these systems would need to be far more advanced and human-like. This is because the model inherently presupposes that the "another" is a human being; someone capable of understanding what is happening, intervening, and acting differently if necessary. Hence, since "tricking" or otherwise seizing control of the intermediary's mental capacity is necessary, the intermediary must possess a certain level of awareness or the capacity to act autonomously, which existing AI systems cannot[716].

It is asserted that the indirect perpetration liability model may be applicable also in German and Turkish legal systems, when AI-driven autonomous systems are utilised in the commission of a crime. Accordingly, robots do not possess culpability of their own; therefore, it may be analogised to a child or a mentally ill person and treated as a tool, with the individual using it to commit a crime being classified as an indirect perpetrator under Article 37(2) of the Turkish Penal Code (TPC)[717] and Section 25(1) of the German Criminal Code (StGB). In this situation, the robot's lack of knowledge regarding the legal context of the offence is being exploited[718].

For a similar perspective, see: DOBRINOIU, The Influence, 2019, p. 144.

715 VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 664.

716 LIMA, Could AI, 2018, p. 690-691.

717 ALTUNÇ, Yapay Zekâ, 2021, p. 354 f.

718 JOERDEN, Strafrechtliche Perspektiven, 2013, pp. 205-206.
See also: MITSCH, Roboter und Notwehr, 2020, p. 372 f

b. Theoretical Basis of Indirect Perpetration

Several theories have been proposed in literature to define the concept of indirect perpetration. While a detailed analysis lies beyond the scope of this study, it can be briefly observed that a common feature in frequently discussed cases of indirect perpetration is the commission of a criminal offence through the use of an intermediary, often described as a human "tool". In such cases, the indirect perpetrator exerts control or dominance over the intermediary's actions, making the intermediary's conduct appear as the work of the person behind[719]. For control to meet the criteria of this model, the person behind must use the person in front (the innocent agent) instrumentally as a tool[720], typically exploiting their lack of rationality or deceiving them[721]. Hence, the actions of the direct perpetrator are attributed to the indirect perpetrator[722].

The concept of control exerted by the person behind over the person in front -despite opposing views- should be interpreted in a normative sense. It pertains to the legal responsibility of the person behind for the legally relevant lack of culpability or responsibility of the person in front, rather than psychological dependencies such as group dynamics, or other forms of influence such as financial dependency. The liability of the person behind is based on any constitutive deficiency in the responsibility of the person in front, whether this deficiency stems from justification, blamelessness, or subjective and/or objective factors[723].

In certain situations, it may be challenging to distinguish between incitement and indirect perpetration. This distinction becomes particularly significant when non-culpable individuals are involved. Under the principle of limited accessory liability, incitement does not require a culpable, but

719 FREUND, § 10 Täterschaft und Teilnahme, 2009, p. 386 Rn. 54.

720 ZIESCHANG, Strafrecht AT, 2023, p. 183 Rn. 664 ff.

721 HORDER, Ashworth's Principles of Criminal Law, 2019, p. 128.

722 The location of the crime is considered both the place of the indirect perpetrator's own activity and the direct perpetrator's act which causes the effects constituting the offence. WERLE/JEßBERGER, § 9 Ort in der Tat in LK, 2020, p. 694, Rn. 14. The time of the crime is considered to be the moment when the direct perpetrator performs the criminal act. DANNECKER/SCHUHR, § 2 Zeitliche Geltung in LK, 2020, p. 375 f., Rn. 46; WERLE/JEßBERGER, § 8 Zeit der Tat in LK, 2020, p. 687, Rn. 9.

723 KINDHÄUSER/ZIMMERMANN, § 39 Alleintäterschaft - Strafrecht AT, 2024, 2024, p. 362 Rn. 8 ff.; SCHÜNEMANN/GRECO, § 25 Täterschaft in LK, 2021, p. 751, 791, Rn. 79, 156.

only an intentional commission of an unlawful offence. Accordingly, under the doctrine of control over the offence, the decisive factor is whether the control over the knowledge or will of the person behind overlaps with the control over the actions of the person in front. If the person behind consciously exploits the lack of culpability or justification of the person in front, the latter is regarded as a tool, and the doctrine of indirect perpetration is applied[724].

The indirect perpetrator's intent must be directed towards fulfilling the objective elements of the offence, encompassing both the knowledge of these elements and the desire to realise them[725]. An indirect perpetrator utilises a person who is unaware that their actions fulfil the objective elements of an offence, meaning they do not realise that their conduct meets the requirements prescribed by criminal law[726]. According to the prevailing opinion and jurisprudence, the criterion is that the indirect perpetrator must have control over the act; mere subordination is not sufficient. The exercise of control may rely on superior knowledge or willpower. The front person's status as a tool arises from their lack of criminal liability, which may stem from a deficiency in the objective elements of the offence, unlawfulness or culpability. This lack of liability may also result from the dominance exerted by the person behind or from the exploitation of an error[727].

c. Assessment

The *ratio legis* of indirect perpetration lies in committing a crime by dominating another's actions through exercising control over their will and using their conduct as a tool to achieve the offence[728]. In light of the aforementioned considerations, the assertion, frequently encountered in literature, that the liability of the person behind arises because the

724 HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 162 Rn. 46; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 186 Rn. 497; JOERDEN, Strafrechtliche Perspektiven, 2013, pp. 205-206

725 HILGENDORF/VALERIUS, Strafrecht AT, 2022, p. 164 Rn. 53.

726 HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 182 Rn. 490.

727 RENGIER, § 43. Mittelbare Täterschaft in Strafrecht AT, 2019, p. 382, 391 Rn. 1 ff., 42; BOHLANDER, Principles of German Criminal Law, 2009, p. 156; KASPAR, § 6 Täterschaft und Teilnahme in Strafrecht AT, 2023, p. 142 f. Rn. 36 ff.; HOFFMANN-HOLLAND, Strafrecht AT, 2015, p. 186 Rn. 498.

728 ÖNOK, Joint Criminal Enterprise, 2019, p. 221.

person in front cannot be punished, requires careful consideration. The non-punishment of the person in front is a consequence, whereas the lack of criminal liability, stemming from specific legal grounds, constitutes the underlying reason. The fundamental premise of criminal law is the imposition of punishment on culpable individuals who fulfil the elements of an offence. It does not seek to attribute liability to another party merely because one individual is exempt from punishment. Therefore, in cases where the person in front is not punished for any reason; whether due to the existence of a personal justification for immunity or otherwise, an issue arises when the person behind is being categorised as an indirect perpetrator when they should actually be considered as an instigator[729].

In my view, considering the current state of technology, applying the indirect perpetration liability model is not only unnecessary but also misguided. First, in the examples provided by *Hallevy*, who advocates for this approach, the focus is not on the autonomous features of AI-driven systems but rather on systems that generate deterministic outputs for a given command. Such systems are merely tools, akin to firearms. When a firearm or a screwdriver is used as a tool to commit a crime, it is classified as a weapon, and the concept of indirect perpetration is not invoked. The opposite way of thinking would imply granting these tools, albeit to a limited extent, a 'will' and the capacity to perform 'acts' in the sense recognised by criminal law; because an innocent agent typically performs the *actus reus* but lacks the requisite *mens rea*[730]. Nonetheless, with due respect to *Hallevy*'s perspective, I find it difficult to accept the notion of attributing "action" to a screwdriver[731]. Additionally, it has been concluded above that AI-driven autonomous systems cannot perform an act in the sense required by criminal law[732]. Furthermore, what is legally challenging is autonomous systems. For example, if programmers of a self-driving vehicle's software deliberately omit any data related to sidewalks during the training phase, with the intent of causing the vehicle to hit pedestrians, this could be considered a genuine instance of the utilisation of an AI-driven autonomous system. In such a case, there would be no need to invoke the concept of indirect perpetration, as direct intentional liability would apply.

729 For the same view, see: ÖNOK, Joint Criminal Enterprise, 2019, p. 224.
Önok provides the example of inciting a member of parliament, who enjoys legislative immunity as a personal ground for exemption from liability, to use profanity during a parliamentary speech.

730 MOLAN/LANSER/BLOY, Principles of Criminal Law, 2000, p. 116.

731 HALLEVY, The Criminal Liability, 2010, p. 180 f.

732 See: Chapter 3, Section B(3): "Can Autonomous Systems 'Act' In the Legal Sense?".

Although robots, while not considered legal persons under criminal law, could, with certain adjustments, be regarded as "will-less tools" and their conduct could be attributed to the human behind them. However, such an approach is unnecessary because current criminal law already allows a robot's conduct to be attributed to the programmer or user through causality, as the programming or deploying serves as the initial trigger. While proving this nexus may be challenging in some instances, existing legal frameworks are deemed sufficient[733].

The desire to apply this model, in my view, stems from a misunderstanding of how AI systems operate. Indeed, this is evidenced by the fact that some scholars propose applying the model to highly autonomous systems (strong AI), while others propose its application solely to low-level autonomous systems (weak AI) conducting entirely under the control of the programmer or operator. Hence, proponents must first address the following question: is this model being applied because AI-driven systems operate autonomously and are therefore analogous to a child, or because they exhibit a slight degree of unpredictability while remaining largely dependent on the person controlling them? It appears that a conceptual inconsistency arises at this point.

What should be highlighted here is that the indirect perpetrator utilises not another person's physical body but their actions as a tool, through exercising control over their will[734]. At the current level of technology, it is not possible to exploit an autonomous system's will through error or to establish dominance over its knowledge or willpower (although manipulating these systems is possible, this does not equate to exercising control over their will)[735]. As for future systems, it is difficult to make a definitive determination at this stage.

According to Section 25(1) of the German Criminal Code (StGB), "[w]hoever commits an offence themselves or through another incurs a penalty as an offender". The understanding currently accepted in both doctrine and legislation is that "another" must be a human being. For instance, when a surgeon uses an AI-driven machine during surgery, the machine cannot be considered "another"; therefore, the surgeon is the sole perpetrator[736]. Similarly, neither animals nor legal persons can be consid-

733 MARKWALDER/SIMMLER, Roboterstrafrecht, 2017, p. 179.
734 ÖNOK, Joint Criminal Enterprise, 2019, p. 221. See also: TÜRAY, Fikir ve Sanat, 2024, p. 624 f.
735 KATOĞLU/ALTUNKAŞ/KIZILIRMAK, Yapay Zekâ, 2025, *passim*.
736 HILGENDORF, Grundfragen, 2013, p. 28.

ered as "another" within the meaning of this provision[737]; therefore, this is even less applicable to AI-driven systems, which lacks personhood and is inherently regarded as merely a tool. The only scenario in which the indirect perpetrator model could be applied is if the manufacturer produces the AI system and makes another person use it to commit a crime. In this case, the innocent agent would be the person operating the AI-driven system[738].

Similar to German Law, Article 37(2) of Turkish Penal Code (TPC)[739] stipulates that "[a]ny person who uses another as an instrument for the commission of an offence shall remain culpable as an offender". Likewise, an equivalent provision exists under U.S. law (18 U.S.C. § 2(b))[740]: "[w]hoever willfully causes an act to be done which, if directly performed by him or another, would be an offense against the United States, is punishable as a principal."

To sum up, considering all of the above, in my view, it is not possible to invoke indirect perpetration in cases where AI-driven autonomous systems are utilised to commit crimes; because: (1) they lack will; (2) their conduct cannot be considered an act in the sense of criminal law, and (3) they are not human to be considered as "another". Even if the requirement for the innocent agent to be human were ignored, and it was accepted that AI-driven autonomous systems could perform acts in the sense of criminal law; they would still need to possess a certain level of will for this debate to hold any meaningful relevance.

3. The Natural Probable Consequence Liability Model

The model proposed by *Hallevy* and widely debated in literature seeks to address the risk of crimes involving AI-driven autonomous systems remaining unpunished, even when such crimes were not directly intend-

737 KINDHÄUSER/ZIMMERMANN, § 39 Alleintäterschaft - Strafrecht AT, 2024, 2024, p. 362 Rn. 7.

738 SCHÄFER, Artificial Intelligence und Strafrecht, 2024, p. 506 f.

739 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).

740 U.S. Department of Justice, "Criminal Resource Manual § 2471: 18 U.S.C. § 2.", https://www.justice.gov/archives/jm/criminal-resource-manual-2471-18-usc-2. (accessed on 01.08.2025).

ed. Accordingly, a crime is initially planned to be committed using an autonomous system, but additional or more severe crimes occur beyond the original intent. It is necessary to ascertain whether the unintended crimes are a natural, probable, and foreseeable consequence of the initially intended act. This would result in the attribution of negligent liability to the programmer. For the model to apply, the unintended crimes must result from the initially planned crime and must have been subjectively foreseeable at the outset. In such cases, the programmer would be held liable for both the intended and unintended crimes. However, if the unintended crime is committed through the influence of an advanced, autonomous AI-driven system, criminal liability may extend to the AI itself, in addition to the programmer. Conversely, if no crime was planned but an autonomous system still causes harm, this model would not apply[741].

In my view, while this model is presented under a different name in the context of Anglo-American legal system, it essentially corresponds to doctrines such as crimes aggravated by their consequences or the liability of accomplices exceeding the scope of the original plan. Ultimately, it does not deviate from the principle of holding the persons behind the machine liable[742]. Indeed, *Hallevy* himself also emphasises that the model is intended to ensure deterrence by encouraging greater caution and diligence among those responsible for AI-driven autonomous systems[743].

741 HALLEVY, The Criminal Liability, 2010, pp. 181-186; HALLEVY, Liability for Crimes Involving AI, 2015, pp. 115-120.

742 For the same view, see: VOJTUS/KORDIK/DRAZOVA, Artificial Intelligence, 2022, p. 664.

743 HALLEVY, Liability for Crimes Involving AI, 2015, p. 119.