Oslo Law Review

SPECIAL ISSUE: AI AND ROBOTICS IN HEALTHCARE
SCIENTIFIC PUBLICATION

Scandinavian University Press

Volume 11, No. 1-2024, p. 1–10
ISSN online: 2387-3299
DOI: https://doi.org/10.18261/olr.11.1.3

AI Systems and Criminal Liability

A Call for Action

Athina Sachoulidou
Assistant Professor in Criminal Law, School of Law, Aristotle University of Thessaloniki; Visiting Senior Researcher, CEDIS, NOVA School of Law, Universidade NOVA de Lisboa
sachoulidou@law.auth.gr

Abstract
The rapid advancement and widespread adoption of artificial intelligence (AI) and other enabling technologies underscores the enduring debate over attributing criminal liability to non-human agents. At the same time, the increasing risks associated with the use of AI systems, which may amount to grave violations of legal interests, such as life, bodily integrity and privacy, raise concerns as to whether one could address AI-related offences by means of employing traditional criminal law categories. In particular, it is questioned whether commonly accepted frameworks rooted in concepts such as personhood, *actus reus*, causation and *mens rea* are adequately equipped to address criminal conduct in AI settings. This article provides an overview of the key points raised as part of this scholarly discourse and presents two primary approaches to criminal liability in the age of AI, namely the use of the 'permissible risk' doctrine and the solution of introducing new endangerment offences, exploring their merits and pitfalls.

Keywords
Artificial intelligence (AI), AI-related offences, criminal liability, endangerment offences, (permissible) risk

1. Introduction

A couple of years ago, one may have found it challenging to depict the increasing impact of artificial intelligence (AI) and other enabling technologies on our daily lives in realistic terms, namely without resorting to the jargon of science fiction movies. Nowadays, automation of this kind is getting a much clearer shape: autonomous vehicles and other driving assistants, chatbots, smart home devices, content generation applications, AI-supported medical diagnostics, and AI-empowered crime detection, to name a few. This inevitably impacts on the lawmaker's tasks, which now include finding ways to address risks emerging from or harms caused by the use of AI systems. At the EU level,[1] there have already been two seminal law-making initiatives seeking to attain this goal: the Proposal for a Regulation laying down harmonised rules on AI (Artificial Intelligence Act; hereinafter AIA),[2] which was

1. For a comprehensive overview of AI law and governance initiatives at international, regional and national level, see Lottie Lane, 'Clarifying Human Rights Standards Through Artificial Intelligence Initiatives' (2022) 71(4) *International & Comparative Law Quarterly* 915 <https://doi.org/10.1017/S0020589322000380> accessed 26 April 2024.
2. European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM/2021/206 final).

formally adopted by the European Parliament on 13 June 2024 and entered into force on 1 August 2024,[3] and the Proposal for a Directive on adapting non-contractual civil liability rules to AI (AI Liability Directive).[4] In particular, the AIA aims 'to improve the functioning of the internal market […] promote the uptake of human centric and trustworthy [AI] while ensuring a high level of protection of health, safety, fundamental rights […], including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems in the Union, and to support innovation'.[5] As such, the prohibitions it introduces and the requirements it sets out for the development of the so-called high-risk AI systems are rather meant to *prevent* risks and harms for the users thereof. The Draft AI Liability Directive is designed to complement them by shifting the focus onto *damage* already caused by an AI system and the need to compensate the injured person in terms of civil law.

The intersection of criminal law and AI systems has also been under scrutiny, but rather as regards the impact of AI on the administration of criminal justice (for instance, the use of AI systems to predict the likelihood of crime commission or to identify actual or future victims of crime).[6] Adapting *criminal* liability to AI has not yet been part of any major legislative initiatives at the EU level, nor in the majority of national legal orders.[7] Yet, this is a topic that has already triggered curiosity among the members of the legal scholarly community that, using the example of autonomous vehicles as an inspiration platform, sought to explore the applicability of traditional criminal law categories to AI systems and AI-related offences and to propose solutions for avoiding impunity when the harmful outcome amounts to a violation of a legal interest.[8]

This article systematises the key challenges identified as part of the aforementioned scholarly discourse, without providing a fully-fledged doctrinal analysis thereof. In particular, it singles out: the lack of legal personhood with respect to the AI system; the difficulties inherent in decoding causation in AI settings; and the current lack of predictability and well-defined standards of care in terms of the major obstacles for ascribing criminal liability on

3. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1.
4. European Commission, Proposal for a Regulation of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) (COM (2022) 496 final).
5. Recital 1 AIA.
6. See, for instance, European Parliament, Resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters (2020/2021(INI)).
7. For an overview of national approaches to criminal liability in AI times see Lorenzo Picotti, 'Traditional Criminal Law Categories and AI: Crisis or Palingenesis?' (2023) 94(1) *Revue Internationale de Droit Pénal* 11. See also the concise overview of solutions proposed in the Singaporean, French and British legal orders in Alice Giannini and Jonathan Kwik, 'Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles' (2023) 34 *Criminal Law Forum* 43, 59ff <https://doi.org/10.1007/s10609-023-09451-1> accessed 26 April 2024.
8. See, for instance, Sabine Gless, Emily Silverman and Thomas Weigend, 'If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability' (2016) 19(3) *New Criminal Law Review* 412 <https://doi.org/10.1525/nclr.2016.19.3.412> accessed 26 April 2024; Dafni Lima, 'Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law' (2018) 69 *South Carolina Law Review* 677; Eric Hilgendorf, 'Dilemma-Probleme beim automatisierten Fahren. Ein Beitrag zum Problem des Verrechnungsverbots im Zeitalter der Digitalisierung' (2018) 130 *Zeitschrift für die gesamte Strafrechtswissenschaft* 674 <https://doi.org/10.1515/zstw-2018-0027> accessed 26 April 2024; Maria Kaiafa-Gbandi, 'Artificial Intelligence as a Challenge for Criminal Law: in Search of a New Model of Criminal Liability' in Susanne Beck, Carsten Kusche and Brian Valerius (eds), *Digitalisierung, Automatisierung, KI und Recht. Festgabe zum 10-jährigen Bestehen der Forschungsstelle RobotRecht* (Nomos 2020) 305–328.

the grounds of negligence (Section 2). Next, this article reflects critically on the two major approaches to criminal liability in AI times: the use of the doctrine of permissible risk and the regulation by means of introducing new (endangerment) offences (Section 3). It concludes with opting for a 'pro regulation' approach.

2. Three Major Challenges Arising from the Attempt to Bridge AI and Criminal Liability

2.1 If not the AI System, Then Who is to Be Held Liable?

'Whoever kills another person […]' serves as a global expression of an *act* that can incur criminal liability. Neither the wishful thinking nor the mere intention to achieve the harmful result are enough for holding *someone* liable in the way that liability is conceived in the realm of criminal law. There may be no universal definition, but there are common denominators of the different approaches to the notion of acting (known as *actus reus*), whether codified[9] or not:[10] some sort of conduct that is either positive (act) or negative (failure to act despite having a legal duty to do so; omission) and that has not only some social relevance but is also the outcome of self-reflection and free will. Against this backdrop, the first struggle the legislator needs to overcome to ascribe *criminal* liability to an AI system is answering the question of whether the latter is capable of acting in the sense of criminal law.[11]

To address this question, one needs to navigate the (blurry) waters of the definition of AI and, particularly, unveil the meaning ascribed to the second component of this term, namely 'intelligence', inasmuch as it is the latter that seems to allow a comparison between humans and AI systems as equal (?) addressees of criminal punishment. The EU legislator has already provided a working definition of AI in Art 3, point (1) AIA[12] that should assist scholars and practitioners in classifying and making better sense out of multiple computational models.[13] Yet, this definition will not necessarily shed any light into the concept of personhood for the purpose of solving the criminal liability knot. At the same time, the existing scholarship suggests that, although further progress is expected in AI settings, AI remains at the moment (and for the foreseeable future) *narrow* (as opposed to general AI). This means that AI systems 'can perform one or few specific tasks'.[14] Building the capabilities required to achieve general AI, including self-awareness and the ability to define its own purpose, is

9. See the definition of criminal liability in the US Model Penal Code (§2.01 (1)): 'A person is not guilty of an offense unless his liability is based on conduct which includes a voluntary act or the omission to perform an act of which he is physically capable.'
10. For a concise presentation of the doctrine of conduct, see Johannes Keiler, 'Actus Reus' in Pedro Caeiro, Sabine Gless, Valsamis Mitsilegas, Miguel João Costa, Janneke De Snaijer and Georgia Theodorakakou (eds), *Elgar Encyclopaedia of Crime and Criminal Justice* (Edward Elgar 2024).
11. Lima (n 8) 680.
12. The definition reads as follows: '"AI system" means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments'.
13. See Chelsea Barabas, 'Beyond Bias: "Ethical AI" in Criminal Law' in Markus Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020) 737.
14. High-Level Expert Group on Artificial Intelligence (AI HLEG), *A definition of AI: Main Capabilities and disciplines* (2018) 6 <https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf> accessed 26 April 2024. Cf Giannini and Kwik (n 7) 47–48, who stress that, even within this limited context, 100% reliability cannot be achieved, as well as that risk of harm arises from, *inter alia*, lack of understandability and bias.

still a technological challenge.[15] This latter note suggests that the discussion on attributing legal personhood or capability to AI systems for the purpose of holding them accountable directly, namely without having to identify the responsible human agent *acting* behind them, is – at least – premature. Even if one considers the bodily dimension of conduct obsolete, particularly in those legal orders where corporate criminal liability is considered a welcome extension of criminal liability beyond human agents, the lack of meaningful purpose and self-awareness contradict the ability for judgement and free will (concepts that continue to pose obstacles to the universal acceptance of criminal liability of legal persons) as a fundamental component of blame and punishment.[16] Thus, in the best-case scenario, we are dealing (again) with a new agent who has *no soul to be damned, and no body to be kicked*[17] – at least, not in a way that would fulfil the purposes of criminal punishment (despite the lack of unanimity as to the prevalent doctrine), ranging from retributivism to rehabilitation.[18]

2.2 Lost in Causation

The rejection of AI personhood does not negate the relevance of AI-related offences, but certainly impacts on the content ascribed to this term. For instance, Hayward and Maas distinguish between:[19] crimes *with* AI, where AI serves as a tool (the same way one uses a weapon or a knife to commit homicide or a computer system to commit fraud); crimes *on* AI, where AI becomes the object of the criminal act (similar to other already established offences such as the attacks against information systems); and crimes *by* AI. In the case of the latter category, where one could subsume car accidents *caused* by autonomous vehicles, AI – should we decline AI personhood – is to be classified as an intermediary the contribution of which to the harmful result is relevant to the extent it can be attributed, whether fully or partially, to a human agent or it can otherwise assist us in evaluating the action (or the lack thereof) of a responsible person.[20] This case appears to be the most problematic considering that, first, the attribution of the AI systems' contribution to a human agent depends on the degree of autonomy of this particular system,[21] and that, second, the expectations we may have regarding the human interaction with AI systems are not, nor should be (based on the current level of development of such systems), unlimited.

This already implies that, besides addressing the question of whether a new agent in the form of an AI system that presents a certain level of autonomy is arising or will arise any time soon, one should evaluate the conduct of a multitude of agents that interact with the AI system and are possibly involved in the commission of AI-related offences.[22] This includes the persons, whether natural or legal, that design, program, train, develop, produce and monitor the AI system, the users and operators thereof as well as those otherwise interacting

15. ibid.
16. See Gless, Silverman Weigend (n 8) 415–417; Lima (n 8) 682–683.
17. This expression is attributed to Edward Thurlow, First Baron Thurlow and Lord Chancellor of England (1731–1806) and has been used extensively in the scholarly discourse on whether legal persons may be held liable in terms of criminal law.
18. For different ways to perceive sanctioning of robots, see Karsten Gaede, *Künstliche Intelligenz – Rechte und Strafen für Roboter?* (Nomos 2019) 99ff.
19. Keith Hayward and Matthijs Maas, 'Artificial intelligence and crime: A primer for criminologists' (2021) 17(2) *Crime, Media, Culture* 209 <https://doi.org/10.1177/1741659020917434> accessed 26 April 2024.
20. See Maria Kaiafa-Gbandi, Athina Sachoulidou and Dafni Lima, 'Greek Report on Traditional Criminal Law Categories and AI' (2023) 94(1) *Revue Internationale de Droit Pénal* 223, 229–232; Susanne Beck and Simon Gerndt, 'German Report on Traditional Criminal Law Categories and AI' (2023) 94(1) *Revue Internationale de Droit Pénal* 195, 202–204.
21. Beck and Gerndt (n 20) 203.
22. See Lima (n 8) 681.

therewith (eg, other drivers in the case of autonomous vehicles or other users of digital platforms in the case of chatbots). Holding these persons liable in the sense of criminal law presupposes that they have not only acted or failed to act in a way that at least absorbs the 'poor judgement' of the AI system (otherwise, one should employ the 'in dubio pro reo' principle),[23] but also there is a *causal link* between this specific conduct (eg, faulty programming, training errors or lack of monitoring)[24] and the harmful result in question.

The analysis of the different outcomes resulting from the adoption of one or another model of causation or causality, including but not limited to the *conditio-sine-qua-non* formula and the doctrine of objective attribution, falls outside the scope of this article.[25] Yet, all different approaches encounter the so-called 'problem of many hands',[26] which is exacerbated in AI settings.[27] Before reaching the market, AI systems are *in principle* developed, trained and produced at industrial scale; that is, there is a multitude of persons, groups and departments that shape the 'developer-manufacturer-user' chain, which is further complemented by those responsible for the post-market monitoring of the product in question.[28] This means that it can be challenging to single out the critical mistake(s) and, in general, to deconstruct the series of interactions of these actors and distinguish the foreseeable from the unforeseeable chain of events as well to evaluate those in the light of the skills and capabilities of each involved party.

This challenge is not new to the criminal-law doctrine, nor unique to AI settings, considering that the complexity of the corporate environment and decision-making has served as one of the main arguments in favour of introducing schemes of corporate criminal liability.[29] Instead, the decisive difference is the *lack of predictability*[30] that is associated with the use of an AI system – often described as the 'black box' that even computer engineers seek to peek inside.[31] In particular, depending on their level of autonomy and learning capabilities, the outcome of using an AI system may differ (substantially) from what the designer or the producer could envision or the (lay) user would expect based on the information provided by the former.[32] This may pose a great obstacle to the verification of the causal link or even negate its existence. In the latter case, a scheme of criminal liability based on over-simplified approaches[33] to causation would violate the principle of guilt[34] – with the human 'in the

23. Kaiafa- Gbandi (n 8) 317; Beck and Gerndt (n 20) 209.
24. Beck and Gerndt (n 20) 209.
25. For different doctrinal approaches to this matter, see Beck and Gerndt (n 20) 209–210; Gless, Silverman and Weigend (n 8) 426–428; Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 238–240.
26. See Dennis Thompson, 'Moral Responsibility and Public Officials: The Problem of Many Hands' (1980) 74(4) *American Political Science Review* 905 <https://doi.org/10.2307/1954312> accessed 26 April 2024.
27. See Giannini and Kwik (n 7) 58–59.
28. After the AI system has been merchandised, one may also have to count with the interconnectedness of different AI systems. See Cornelius Kai, 'Künstliche Intelligenz, Compliance und sanktionsrechtliche Verantwortlichkeit (2020) *Zeitschrift für Internationale Strafrechtsdogmatik* 51, 55.
29. For an analysis of the role of corporate structure and organisation as hurdles associated with the attribution of criminal liability in criminal settings see Athina Sachoulidou, *Unternehmensverantwortlichkeit und -sanktionierung. Ein strafrechtlicher und interdisziplinärer Diskurs* (Mohr Siebeck 2019) 60–65.
30. See Ugo Pagallo, 'When Morals Ain't Enough: Robots, Ethics and the Rules of Law' (2017) 27 *Minds and Machines* 625 <https://doi.org/10.1007/s11023-017-9418-5> accessed 26 April 2024.
31. For the concept of explainable AI (known as XAI), see Amina Adadi and Mohammed Berrada, 'Peeking Inside the Black-Box: a Survey on Explainable Artificial Intelligence (XAI)' (2018) *IEEE Access* 6, 52138 <https://doi.org/10.1109/ACCESS.2018.2870052> accessed 26 April 2024.
32. Similarly, Giannini and Kwik (n 7) 52, 54, who also underline the difference between rule-based AI and AI based on machine learning as well as that the users may feel demotivated from acquiring more knowledge regarding the AI systems they deploy if this knowledge could facilitate the establishment of their liability.
33. Cf Giannini and Kwik (n 7) 45, who argue in favour of 'more refined regimes or interpretations of law' in order to 'avoid unequitable attribution of responsibility or scapegoating'.
34. Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 240

loop' being treated as a mere object of liability[35] or as a mere means to achieve the end of reducing impunity in AI settings. Thus, acknowledging that there may be a gap of responsibility[36] in such cases would be an optimal solution that also complies with the fragmentary character of criminal law.

2.3 Mens Rea without Boundaries?

Shifting the focus onto *mens rea*, many authors argue that the cases where one deploys an AI system as an instrument with the intention to cause harm (eg, violations of property, privacy and data protection or even human life)[37] are less problematic or may present difficulties already addressed in the criminal-law doctrine (eg, the crime committed with the use of the AI system goes beyond the original intent).[38] This is only partially true inasmuch as intention, whether in the form of *dolus directus* (first or second degree) or in the form of *dolus eventualis*, consists of a *cognitive* (knowledge) and a *voluntative* element (will). To ascribe liability for an intentional offence, the offender should at least be able to objectively foresee that his/her conduct is likely to cause the harmful effect or the danger prescribed in law, but proceeds to act or omits to do so regardless of the consequences of his/her conduct. In AI settings, however, one may not comprehensively understand the inner workings of the system (s)he employs or what the consequences of the system's decision will be and, thus, there can be no certainty or even strong likelihood that the AI system will serve his/her criminal intent. As Beck points out, there may be an abstract perception of the possibility of causing a damage, but it remains doubtful whether this generic knowledge is enough to establish liability on the grounds of intent.[39]

Negligent offences[40] are expected to absorb the majority of criminal incidents involving AI systems, but the establishment of negligence can be equally challenging. Compared to intent, negligence stands for a qualitatively different combination of knowledge and will. Again, the offender needs to be able to at least foresee that his/her conduct could cause a specific result, although (s)he does not accept its occurrence and may have wished to avoid it. Even, in the case of the so-called 'unconscious negligence', where one does not foresee the harmful consequences of his/her behaviour and, thus, there can be no kind of will attached to them, criminal liability will be ultimately linked to the capability of this specific person to foresee the violation of the legal interest in question under the concrete circumstances.[41] This means that negligence cannot, nor should, serve as a gatekeeper of criminal liability in those cases where the autonomous learning and development (information analysis and

35. Susanne Beck, 'Digitalisierung und Schuld' in Thomas Fischer and Elisa Hoven (eds), *Schuld*, Band 3 (Nomos 2017) 289, 298.
36. See Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 *Ethics and Information Technology* 175–183 <https://doi.org/10.1007/s10676-004-3422-1> accessed 26 April 2024.
37. Cf Thomas King, Nikita Aggarwal, Mariarosaria Taddeo and Luciano Floridi, 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions' (2020) 26 *Science and Engineering Ethics* 89 <https://doi.org/10.1007/s11948-018-00081-0> accessed 26 April 2024.
38. For instance, Lima (n 8) 690–691; Daniele Amoroso and Benedetta Giordano, 'Who Is to Blame for Autonomous Weapons Systems' Misdoings?' in Elena Carpanelli and Nicole Lazzerini (eds), *Use and Misuse of New Technologies* (Springer 2019) 217; King and others (n 37) 109; Giannini and Kwik (n 7) 48–49.
39. See Susanne Beck, '§7' in Martin Ebers, Christian Heinze, Tina Krüge; and Björn Steinrötter (eds), *Rechtshandbuch – Künstliche Intelligenz und Robotik* (Beck 2020) 249. Similarly, Giannini and Kwik (n 7) 55.
40. For the purposes of this analysis, negligence is only perceived as a form of *mens rea*. For an analysis of the parameter of 'objective negligence' as a fundamental component of negligent AI-related offences, see Kaiafa-Gbandi (n 8) 314–319; Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 236–238.
41. Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 230.

response) of the AI system negates the predictability of the harmful event for those designing, developing, manufacturing, monitoring or using the system. There can be exceptions to this rule. As Kaiafa-Gbandi correctly notes, liability should not be waived when the individual takes over the role of the system's supervisor in terms of monitoring AI learning with the aim of intervening and interrupting the use of the system in place in case of dangerous system discrepancies, provided such a duty is prescribed in law.[42] Even this solution does not necessarily guarantee better results in terms of establishing liability, unless the supervisor 'has actual *meaningful* control over subsequent events', a capability that may be impacted adversely due to passive monitoring (as the example of autopilot has proved) or lack of time.[43]

Lastly, no matter if one places the notion of negligence exclusively in the area of *mens rea* or ascribes thereto additional objective dimensions (an analysis that is beyond the scope of this article),[44] the *standard of due care* (otherwise the duty to take appropriate and reasonable care and, in particular, the failure to exercise it) is of central importance for establishing criminal liability on the grounds of negligence. This suggests that there should be different expectations towards each person involved in the design, development, training, production, monitoring or use of an AI system depending on their exact role and tasks. Furthermore, in certain cases, national or supranational (hard) laws or even non-governmental standards (eg, ISOs) may stipulate positive obligations for this purpose. The future implementation of the AIA is expected to signal the proliferation of such rules and standards, the number of which remains for the moment rather limited.[45] Yet, the involvement of the private sector in this rule-making procedure raises concerns as to whether these requirements will be shaped with the aim of ensuring the maximum care for the legal interests at stake or with the aim of minimising the financial burden on the respective industry.[46] In any event, they will serve as a basis for substantiating the claim of lawfulness of the conduct of AI designers, developers, manufacturers, users etc[47] and thus for negating, *inter alia*, criminal liability on the grounds of negligence.

3. Two (Imperfect) Solutions at the Disposal of the (Supra-)National Legislator

3.1 Permissible Risk

The risks that arise in AI environments serve as a reminder of the challenges that technological progress, such as the mass-scale production and use of cars, the function of industrial plants operating with nuclear energy, or the globalisation of the internet, has historically presented to legislators, as well as of the fruitless nature of general or absolute prohibitions on endangerment that would defeat the social benefit arising from this progress.[48] In this

42. Kaiafa-Gbandi (n 8) 323.
43. Giannini and Kwik (n 7) 57.
44. See Maria Kaiafa-Gbandi, *Objective and subjective negligence in criminal law* [in Greek] (Sakkoulas 1994).
45. See Georg Borges, 'Rechtliche Rahmenbedingungen für autonome Systeme' (2018) *Neue Juristische Wochenschrift* 977–982; Beck and Gerndt (n 20) 212.
46. Regarding the shortcomings associated with the standardisation model promoted by the AIA, see Martin Ebers, 'Standardizing AI. The Case of the European Commission's Proposal for an "Artificial Intelligence Act"' in Larry DiMatteo, Cristina Poncibò and Michel Cannarsa (eds) *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 321–344.
47. Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 237.
48. Detlev Sternberg-Lieben and Frank Peter Schuster, '§15' in *Schönke/Schröder Strafgesetzbuch* (30th edn, CH Beck 2019) para 144. As to the criterion of social benefit, see Cornelius Prittwitz, *Strafrecht und Risiko: Untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft* (Klostermann 1993) 227ff.

context, the doctrine of permissible risk (also referred to as socially acceptable risk)[49] is used to cap the aforementioned standards of due care and, thus, to limit the scope of the *neminem laedere* principle.[50] This suggests that the legal order *and the victim* of an (undesirable) criminal act should accept a minimal level of risk that 'persists' even when all precautions are taken, namely when the duty to take *appropriate* and *reasonable* care is properly exercised. Such tolerance is expected in the light of the permission granted by the same legal order to certain activities due to the societal benefit associated therewith. This stance should be re-examined if it becomes evident that the trust in the effectiveness of the care obligations designed to mitigate the risk is not or no longer justified.[51] This implies that this 'gap of responsibility by design' is accepted only to the extent that the current state-of-the-art is closely observed. In this context, ascribing the duty of after-sale monitoring to the manufacturer of the AI system – compared to his/her privileged position to predict and mitigate risks compared to the end users – could be considered, particularly for those cases where the end-product is subject to updates and patches.[52] Besides this, generally foreseeable product discrepancies will not justify the establishment of criminal liability, unless they amount to design, development, training, production, monitoring or operation-related faults that could have been prevented in the first place.

The use of the 'permissible risk' doctrine is not merely of theoretical interest. It can have practical implications inasmuch as it will mean that the victim will have to tolerate certain harms without access to redress. Besides this, it has substantial societal implications, considering that it encompasses an answer to the question of how much risk our societies can and should tolerate in the age of AI,[53] *but* at a moment when neither the use of AI systems nor the risks emerging therefrom have *yet* been normalised (as this may be the case once they 'intrude' into the global market and their use becomes regular). This makes it notably challenging for the lawmaker to establish accurately the initial threshold for impermissible risk, as (s)he has to strike a satisfying balance between fostering innovation and ensuring safety, a task that one should not underestimate, as the lengthy negotiations on the AIA have showcased. This balancing act may remain a crucial aspect of the ever-evolving process of law-making, but it seems that the rapid pace of AI advancement is leaving the legislator struggling to keep up.

3.2 New Endangerment Offences

The regulation of (im)permissible risk in AI settings can take place either outside or *within* the realm of criminal law. Considering the gap of responsibility that may arise due to the difficulties inherent in identifying the responsible human agent, the acts of whom one can evaluate in terms of criminal law, establishing the causal link between his/her conduct and

49. For a comprehensive overview of the different views on the function of the doctrine of the socially acceptable risk (exclusion or justification of wrongfulness), see Maria Kaiafa-Gbandi, Athina Sachoulidou and Nikolaos Chatzinikolaou, 'Offences of Endangerment (Actual and Abstract Danger)' in Pedro Caeiro, Sabine Gless, Valsamis Mitsilegas, Miguel João Costa, Janneke De Snaijer and Georgia Theodorakakou (eds), *Elgar Encyclopaedia of Crime and Criminal Justice* (Edward Elgar 2024) (forthcoming).
50. Sternberg-Lieben and Schuster (n 48) para 145.
51. ibid, paras 145–147.
52. Eric Hilgendorf, 'Autonome Systeme, künstliche Intelligenz und Roboter' in Stephan Barton and others (eds), *Festschrift für Thomas Fischer* (CH Beck 2018) 99, 105–106, 109.
53. See Susanne Beck, 'Intelligent Agents and Criminal Law – Negligence, Diffusion of Liability and Electronic Personhood' (2016) 86 *Robotics and Autonomous Systems* 138, 141; Prittwitz (n 48) 297, 298; Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 240.

the harmful event and, lastly, overcoming the *mens rea* limitations, one may choose to intervene by means of expanding criminalisation and, particularly, adopting new (abstract) endangerment offences that will compensate for the lack of predictability that, based on the previous analysis, seems to be a horizontal obstacle.[54] Such an intervention may mean the criminalisation of the establishment or maintenance of a functional source of danger in the form of releasing an AI system with unlimited learning possibilities and without sufficient safeguards.[55]

This solution would release the judiciary from the burden of proving causation, but it remains doubtful whether it would foster innovation or, instead, stall the development of learning capabilities of AI systems, a development that is considered *prima facie* beneficial to the extent that it allows the system to improve its performance over time (despite the unpredictability it goes together with).[56] Besides this, this may pave the way to over-regulation in the form of over-criminalisation, an outcome that would be hardly compatible with the ancillary nature of criminal law and the protection of fundamental rights in this realm.[57]

Lastly, it is equally important to address the question of *who* may adopt these new AI-related endangerment offences. In the EU ecosystem, one would expect the national legislator to take the lead, considering that this step may presuppose interventions in the area of the General Part of Criminal Law, the provisions of which are not harmonised and which, given the cultural differences of the European legal orders, should preserve its unique character to avoid conflicts with fundamental tenets of national criminal justice systems.[58] This approach underestimates, however, the fact that the majority of EU Member States are not producers of AI systems, nor are systematically involved in the development thereof. This means that the know-how required to proceed with criminalisation of this kind will most likely be missing at national level.[59]

On the other hand, resorting to the mechanisms of harmonisation of substantive criminal law as set out in Article 83 TFEU exposes us to their inherent limitations, whether one seeks to meet the criterion of serious, cross-border crime (para 1) in order to justify the expansion of the list of the so-called euro-crimes or resorts to the solution of 'functional' substantive criminal law (para 2).[60] The latter solution presupposes an existing Union policy in an area that has already been subject to harmonisation measures.[61] In this case, the harmonisation of criminal laws should assist in securing the effective implementation of this policy, while a more cautious approach (that complies with the principle of subsidiarity and proportionality) would require *a serious enforcement deficit* that can only be resolved through the threat of punishment.[62] This conclusion would be – maybe with the exception of autonomous vehicles, the production of which has already been subject to harmonisation of a certain level[63] – rather premature for the majority of AI systems.

---

54. Hilgendorf (n 52) 110–111.
55. ibid; Kaiafa-Gbandi, Sachoulidou and Chatzinikolaou (n 49).
56. Cf Giannini and Kwik (n 7) 52.
57. Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 240.
58. Cf Art 83(3) TFEU.
59. Similarly, Beck and Gerndt (n 20) 218 who note that 'national solo efforts concerning AI are not to be expected in many spheres of life'.
60. Cf Valsamis Mitsilegas, *EU Criminal Law* (2nd edn, Hart Publishing 2022) 47ff.
61. See Sabine Gless and Katalin Ligeti, 'Regulating Driving Automation in the European Union – Criminal Liability on the Road Ahead?' (2023) *15(1) New Journal of European Criminal Law* 33, <https://doi.org/10.1177/20322844231213336> accessed 26 April 2024.
62. See BVerfG, Judgment of the Second Senate of 30 June 2009 – 2 BvE 2/08, paras 361ff.
63. See Gless and Ligeti (n 61).

4. Conclusion

As the development of AI systems progresses, legislative inaction leaves violations of fundamental rights, which has already taken the form of criminalisation in other areas of activity, without an answer, denying victims access to criminal-law remedies.[64] The implementation of the AIA marks a major step towards regulation in this field, impacting on decisions related to the definition of (un)acceptable risk. The subsequent intervention of criminal law is crucial in order to maintain public trust in the validity of standards as well as to reduce legal uncertainty for the AI industry's stakeholders.[65] As part of this intervention, criminalisation presupposes a thorough understanding of the nuances of this industry and, thus, should proceed on an interdisciplinary ground.

64. Cf Kaiafa-Gbandi, Sachoulidou and Lima (n 20) 247.
65. Beck and Gerndt (n 20) 218.