4

Forms of Robot Liability

Criminal Robots and Corporate Criminal Responsibility

THOMAS WEIGEND

I The Responsibility Gap

The use of artificial intelligence (AI) makes our lives easier in many ways. Search engines, driver's assistance systems in cars, and robots that clean the house on their own are just three examples of devices that we have become reliant on, and there will undoubtedly be many more variants of AI accompanying us in our daily lives in the near future. Yet, these normally benevolent AI-driven devices can suddenly turn into dangerous instruments: self-driving cars may cause fatal accidents, navigation software may mislead human drivers and land them in dangerous situations, and a household robot may leave the home on its own and create risks for pedestrians and drivers on the street. One cannot help but agree with the pessimistic prediction that "[a]s robotics and artificial intelligence (AI) systems increasingly integrate into our society, they will do bad things."[1] If a robot's[2] malfunctioning can be proved to be the result of inadequate programming[3] or testing, civil and even criminal liability of the human being responsible for manufacturing or controlling the device can provide an adequate solution – *if* it is possible to identify an individual who can be blamed for being reckless or negligent in producing, coding, or training the robot.

[1] Mark A. Lemley & Bryan Casey, "Remedies for Robots" (2019) 86:5 *University of Chicago Law Review* 1311 ["Remedies for Robots"] at 1313. For a brief overview of applications of AI and the legal issues related to them, see Eric Hilgendorf, "Modern Technology and Legal Compliance" in Eric Hilgendorf & Maria Kaiafa-Gbandi (eds.), *Compliance Measures and Their Role in Greek and German Law* (Athens: Π.Ν. ΣΑΚΚΟΥΛΑΣ, 2017) 21 at 27–33. For problems associated with controlling self-driving cars, see Chapter 15 in this volume.

[2] Although I am aware that the terms "AI device" and "robot" have slightly different connotations, I use them interchangeably in this chapter.

[3] On the liability of programmers, see Chapter 2 in this volume.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

But two factors make it unlikely that an AI device's harmful action can always be traced back to the fault of an individual human actor. First, many persons, often belonging to different entities, contribute to getting the final product ready for action; if something goes wrong, it is difficult to even identify the source of malfunctioning, let alone an individual who culpably caused the defect. Second, many AI devices are designed to learn from experience and to optimize their ability to reach the goals set for them by collecting data and drawing "their own conclusions."[4] This self-teaching function of AI devices greatly enhances their functionality, but also turns them, at least to some extent, into black boxes whose decision-making and actions can be neither predicted nor completely explained after the fact. Robots can react in unforeseeable ways, even if their human manufacturers and handlers did everything they could to avoid harm.[5] It can be argued that putting a device into the hands of the public without being able to predict exactly how it will perform constitutes a basis for liability, but among other issues it is not clear whether this liability ought to be criminal liability.

This chapter considers two novel ways of imposing liability for harm caused by robots: holding robots themselves responsible for their actions, and corporate criminal responsibility (CCR). It will be argued that it is at present neither conceptually coherent nor practically feasible to subject robots to criminal punishment, but that it is in principle possible to extend the scope of corporate responsibility, including criminal responsibility if recognized in the relevant jurisdiction, to harm caused by robots controlled by corporations and operating for their benefit.

II Robots as Criminals?

To resolve the perceived responsibility gap in the operation of robots, one suggestion has been to grant legal personhood to AI devices, which could make them liable for the harm they bring about. The issue of recognizing

[4] For an interesting example of the logical but dysfunctional learning process of a drone, see "Remedies for Robots", note 1 above, at 1313: A drone was trained to stay within a certain circle and to head toward the center. If the drone left the circle, it was shut off and someone picked it up on the ground and carried it back into the circle. The drone thus "learned" to leave the circle whenever it got close to the margin, because it could then rely on being carried back into the circle.

[5] See Mihailis E. Diamantis, "Algorithms Acting Badly: A Solution from Corporate Law" (2021) 89:4 *George Washington Law Review* 801 ["Algorithms Acting Badly"] at 821–822; Sabine Gless, Emily Silverman, & Thomas Weigend, "If Robots Cause Harm, Who Is to Blame?" (2016) 19:3 *New Criminal Law Review* 415 ["If Robots Cause Harm"] at 426–428.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

E-persons was discussed within the European Union when the European Parliament presented this option.[6] The idea has not been taken up, however, in the EU Commission's 2021 Proposal for an Artificial Intelligence Act,[7] which mainly relies on strictly regulating the marketing of certain AI devices and holding manufacturers and users responsible for harm caused by them. Although the notion of imprisoning, fining, or otherwise punishing AI devices must appear futuristic,[8] some scholars favor the idea of extending criminal liability to robots, and the debate about this idea has reached a high intellectual level.[9] According to recent empirical research, the notion of punishing robots is supported by a fairly large percentage of the general population, even though many people are aware that the normal purposes of punishment cannot be achieved with regard to AI devices.[10]

II.A Approximating the Responsibilities of Machines and Legal Persons

As robots can be made to look and act more and more like humans, the idea of approximating their movements to human acts becomes more plausible – which might pave the way to attributing the notion of *actus*

[6] European Union, European Parliament, Committee on Legal Affairs, Report with Recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL) (Strasbourg, France: European Parliament, January 27, 2017) at 8, www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.pdf. For a brief account of the ensuing discussion, see Anat Lior, "AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy" (2020) 46:5 *Mitchell Hamline Law Review* 1043 ["AI Entities"] at 1067–1069. See also Roman I. Dremliuga, Alexey Yu Mamychev, O. A. Dremliuga *et al.*, "Artificial Intelligence as a Subject of Law: Pros and Cons" (2019) VII:1 *Revista Dilemas Contemporáneos: Educación, Política y Valores* 1 at 9–12.

[7] European Union, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts, COM/2021/206 final (Brussels, Belgium: European Commission, April 21, 2021).

[8] See e.g., "Algorithms Acting Badly", note 5 above, at 807; "AI Entities", note 6 above, at 1070–1071.

[9] See Ying Hu, "Robot Criminals" (2019) 52:2 *Michigan Journal of Law Reform* 487 at 491; Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Cham, Switzerland: Springer, 2015); Gabriel Hallevy, "The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control" (2010) 4:2 *Akron Intellectual Property Journal* 171. For a discussion, see "If Robots Cause Harm", note 5 above, at 415–422.

[10] Gabriel Lima, Meeyoung Cha, Chihyung Jeon *et al.*, "The Conflict between People's Urge to Punish AI and Legal Systems" (2021) 8 *Frontiers in Robotics and AI* Article 756242.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

reus to robots' activities. By the same token, robots' ways of processing information and turning it into a motive for getting active may approach the notion of *mens rea*. The law might, as Ryan Abbott and Alex Sarch have argued, "deem some AIs to possess the functional equivalent of sufficient reasoning and decision-making abilities to manifest insufficient regard" of others' protected interests.[11]

Probably the most sophisticated argument to date in favor of robots' criminal responsibility has been advanced by Monika Simmler and Nora Markwalder.[12] These authors reject as ideologically based any link between the recognition of human free will and the ascription of culpability;[13] they instead subscribe to a strictly functionalist theory of criminal law that bases criminal responsibility on an "attribution of freedom as a social fact."[14] In such a system, the law is free to "adopt a concept of personhood that depends on the respective agent's capacity to disappoint normative expectations."[15] The essential question then becomes "whether robots can destabilize norms due to the capacities attributed to them and due to their personhood and if they produce a conflict that requires a reaction of criminal law."[16] The authors think that this is a probable scenario in a foreseeable future: robots could be "experienced as 'equals' in the sense that they are constituted as addressees of normative expectations in social interaction like humans or corporate entities are today."[17] It would then be a secondary question in what symbolic way society's disapproval of robots' acts were to be expressed. It might well make sense to *convict* an AI device of a crime – even if it lacks the sensory, intellectual, and moral sensibility of feeling the impact of any traditional punishment.[18] Since the future is notoriously difficult to foresee, this concept of robots' criminal responsibility can hardly be disproved, however unlikely it may appear today that humans could have normative expectations of robots and

[11] Ryan Abbott & Alex Sarch, "Punishing Artificial Intelligence: Legal Fiction or Science Fiction" (2019) 53:1 *UC Davis Law Review* 323 ["Punishing Artificial Intelligence"] at 357.

[12] Monika Simmler & Nora Markwalder, "Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence" (2019) 30:1 *Criminal Law Forum* 1 ["Guilty Robots"].

[13] Ibid. at 16: "Idealistic philosophy cannot obscure the fact that the attribution of capacity to reflect, of consciousness, and of other capacities is just that – an attribution – and not cognizable and legally meaningful due to ontological circumstances."

[14] Ibid. at 15.

[15] Ibid. at 17.

[16] Ibid. at 25.

[17] Ibid. at 30.

[18] Cf. "Punishing Artificial Intelligence", note 11 above, at 365–367.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

that disappointment of these expectations would call for the imposition of sanctions. However, in the brave new functional world envisioned by these authors, the term "criminal sanctions" appears rather old-fashioned, because it relies on concepts more relevant to human beings, such as censure, moral blame, and retribution (see Section II.B).

One recurring argument in favor of imposing criminal responsibility on AI devices is the asserted parallel to the criminal responsibility of corporations (CCR).[19] CCR will be discussed in more detail in the following section of this chapter, but it is addressed briefly here because calls for the criminal responsibility of corporations and of robots are reactions to a similar dilemma. In each case, it is difficult to trace responsibility for causing harm to an individual person. If, e.g., cars produced by a large manufacturing firm are defective and cause fatal accidents, it is safe to say that something must have gone wrong in the processes of designing, testing, or manufacturing the relevant type of car. But it may be impossible to identify the person(s) responsible for causing the defect, especially since the companies involved are unlikely to actively assist in the police investigation of the case. As we have seen, harm caused by robots leads to similar problems concerning the identification of responsible humans in the background. Regarding commercial firms, the introduction of CCR, which has spread from the United States to many other jurisdictions,[20] has helped to resolve the problem of the diffusion of responsibility by making corporations criminally liable for any fault of their officers or even – under the *respondeat superior* doctrine – of their employees. The main goals of CCR are to obtain redress for victims and give corporations a strong incentive to improve their compliance with relevant legal rules. If criminal liability is imposed on the corporation whenever it can be proved that one of its employees must have caused the harm, it can be expected that corporations will do everything in their power to properly select, train, and supervise their personnel. The legal trick that leads to this desired result is to treat corporations as or like responsible subjects under criminal law, even though everyone knows that a corporation is a mere product of legal rules and therefore cannot physically act, cannot form an intent,

[19] See e.g., Federico Mazzacuva, "The Impact of AI on Corporate Criminal Liability: Algorithmic Misconduct in the Prism of Derivative and Holistic Theories" (2021) 92:1 *Revue Internationale de Droit Pénal* 143 ["Impact of AI"] at 146–147; "Punishing Artificial Intelligence", note 11 above, at 357; "Guilty Robots", note 12 above, at 18–19 and 27–28.

[20] For a comparative overview, see Francisco Javier Bedecarratz Scholz, *Rechtsvergleichende Studien zur Strafbarkeit juristischer Personen* (Comparative Studies on the Punishability of Legal Persons) (Zurich, Switzerland: Dike Verlag (in cooperation with Nomos), 2016).

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

and cannot understand what it means to be punished. If applying this fiction to corporations has beneficial effects,[21] why should this approach not be used for robots as well?

II.B Critical Differences

However attractive that idea sounds, one cannot help but note that there exist significant differences between corporations and AI devices. Regarding the basic requirements of criminal responsibility, robots at their present stage of development cannot make free decisions, whereas corporations can do so through their statutory organs.[22] At the level of sanctioning, corporations can – through their management – be deterred from committing further offenses, they can compensate victims, and they can improve their operation and become better corporate citizens. Robots have none of these abilities,[23] although it is conceivable that their performance can be improved through reprogramming, retraining, and special supervision. The imposition of retributive criminal sanctions on robots would presuppose, however, that they can in some way feel punished and can link the consequences visited upon them to some prior malfeasance on their part. Today's robots lack this key feature of punishability, although their grandchildren may well be imbued with the required sensitivity to moral blame.

The differences between legal persons and robots do not necessarily preclude the future possibility of treating robots as criminal offenders. But the fact that corporations, although they are not human beings, can be recognized as subjects of the criminal law does not per se lend sufficient plausibility to the idea of granting the same status to today's robots.

There may, however, be another way of establishing criminal responsibility for robots' harmful actions: corporations that use AI devices and/or benefit from their services could be held responsible for the harm they cause. To make this argument, one would have to show that: (1) corporate responsibility as such is a legitimate feature of the law; and (2) corporations can be held responsible for robots as well as for their human agents.

[21] For counterarguments, see text on notes 28–32 below.

[22] Nora Osmani, "The Complexity of Criminal Liability of AI Systems" (2020) 14:1 *Masaryk University Journal of Law and Technology* 53 ["Criminal Liability of AI"] at 61; Dafni Lima, "Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law" (2018) 69:3 *South Carolina Law Review* 677 ["AI Agents"] at 682–683.

[23] Vikram R. Bhargava & Manuel Velasquez, "Is Corporate Responsibility Relevant to Artificial Intelligence Responsibility?" (2019) 17:3 *Georgetown Journal of Law and Public Policy* 829 at 836.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

III Corporate Criminal Responsibility for Robots

III.A Should There Be Corporate Criminal Responsibility?

Before we investigate this option, we should reflect on the legitimacy of the general concept of CCR. If that concept is ethically or legally doubtful or even indefensible, we should certainly refrain from extending its reach from holding corporations responsible for the acts of their human employees to holding them responsible for their robots.

Two sets of theories have been developed for justifying the imposition of criminal responsibility of legal persons for the harmful acts of their managers and employees. One approach regards certain decision-makers within the corporation as its alter ego and therefore proposes that acts of these persons are attributed to the corporation; the other approach targets the corporation itself and bases its responsibility on its criminogenic or improper self-organization.[24] These two theories are not mutually exclusive. For example, Austrian law combines both approaches: its statute on the responsibility of corporations imposes criminal liability on a corporation if a member of its management or its control board committed a criminal offense on the corporation's behalf or in violation of its obligations, or if an employee unlawfully committed a criminal offense and the management could have prevented or rendered significantly more difficult the perpetration by applying due diligence.[25]

Whereas in the United States CCR has been recognized for more than a century,[26] its acceptance in Europe has been more hesitant.[27] In Germany, a draft law on corporate responsibility with semi-criminal

[24] For an overview, see Celia Wells, "Corporate Criminal Responsibility" in Stephen Tully (ed.), *Research Handbook on Corporate Legal Responsibility* (Cheltenham, UK: Edward Elgar, 2005) 147.

[25] *Verbandsverantwortlichkeitsgesetz* (Corporate Responsibility Act), Austria (as amended on May 20, 2016), § 3.

[26] The seminal Supreme Court decision in favor of CCR was *New York Central & Hudson River Railroad Co.* v. *United States*, 212 U.S. 481 (1909). "Algorithms Acting Badly", note 5 above, at 817, correctly observes that today there is great public support in the United States for a broad version of CCR, so that an effort at legislative reform would be a "non-starter." For a report on the present practice of CCR in the United States, see Elisa Hoven & Thomas Weigend, "Praxis und Probleme des Verbandsstrafrechts in den USA" (Practice and Problems of Corporate Criminal Liability in the US) (2018) 130:1 *Zeitschrift für die gesamte Strafrechtswissenschaft* 213.

[27] For a brief overview, see Bernd Schünemann & Luis Greco, "Vorbemerkungen zu §§ 25 para 21" in Gabriele Cirener, Henning Radtke, Ruth Rissing-van Saan *et al.* (eds.), *Strafgesetzbuch. Leipziger Kommentar* (Penal Code, Leipzig Commentary), vol. 2, 13th ed. (Berlin, Germany: De Gruyter, 2021).

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

features failed in 2021 due to internal dissent within the coalition government of the time.[28] Critics claim that CCR violates fundamental principles of criminal law.[29] They maintain that a corporation cannot be a subject of criminal law because it can neither act nor make moral judgments.[30] Moreover, a fine imposed on a corporation is said to be unfair because it does not punish the corporation itself, but its shareholders, creditors, and employees, who cannot be blamed for the faults of managers.[31]

It can hardly be denied that CCR is a product of crime-preventive pragmatism rather than of theoretically consistent legal thinking. The attribution of managers' and/or employees' harmful acts to the corporation, cloaked with sham historical dignity by the Latin phrase *respondeat superior*, is difficult to justify because it leads to a duplication of responsibility for the same crime.[32] It is doubtful, moreover, whether the moral blame

[28] See Germany, Bundesrat, Entwurf eines Gesetzes zur Stärkung der Integrität in der Wirtschaft (Draft Law on the Strengthening of Integrity in the Economy), Bundesratsdrucksache 440/20 (Germany: Bundesrat, August 7, 2020). The draft was not voted on before the parliamentary period ended in the fall of 2021.

[29] For critical assessments, see Ulfrid Neumann, "Zur (Un)Vereinbarkeit des Verbandsstrafrechts mit Grundprinzipien des tradierten Individualstrafrechts" (On the (In-)Compatibility of Corporate Criminal Law with Basic Principles of Traditional Criminal Law for Individuals) in Marianne Johanna Lehmkuhl & Wolfgang Wohlers (eds.), *Unternehmensstrafrecht* (Basel, Switzerland: Helbing Lichtenhahn Verlag, 2020) 49; Frauke Rostalski, "Neben der Spur: Verbandssanktionengesetzgebung auf Abwegen" (Off the Track: Legislation on Corporate Criminal Liability Going Off the Road) (2020) 73:29 *Neue Juristische Wochenschrift* 2087; Uwe Murmann, "Unternehmensstrafrecht" (Corporate Criminal Law) in Kai Ambos & Stefanie Bock (eds.), *Aktuelle und grundsätzliche Fragen des Wirtschaftsstrafrechts* (Berlin, Germany: Duncker & Humblot, 2019) 57; Franziska Mulch, *Strafe und andere staatliche Maßnahmen gegenüber juristischen Personen* (Punishment and Other State Measures against Legal Persons) (Berlin, Germany: Duncker & Humblot, 2017); Friedrich von Freier, "Zurück hinter die Aufklärung: Zur Wiedereinführung von Verbandsstrafen" (Back Behind Enlightenment: On the Re-Introduction of Criminal Punishment for Corporations) (2009) 156 *Goltdammer's Archiv für Strafrecht* 98; Arbeitsgruppe Strafbarkeit juristischer Personen, "Bericht" (Working Group Punishability of Legal Persons, "Report") in Michael Hettinger (ed.), *Reform des Sanktionenrechts*, vol. 3 (Baden-Baden, Germany: Nomos, 2002) 7. For an overview of the recent German discussion, see Thomas Weigend, "Corporate Responsibility in Germany" in Khalid Ghanayem & Yuval Shany (eds.), *The Quest for Core Values in the Application of Legal Norms: Essays in Honor of Mordechai Kremnitzer* (Cham, Switzerland: Springer, 2021) 103.

[30] "AI Agents", note 22 above, at 688.

[31] Mihailis E. Diamantis, "The Law's Missing Account of Corporate Character" (2019) 17:3 *Georgetown Journal of Law and Public Policy* 865 at 880.

[32] See Charlotte Schmitt-Leonardy, "Originäre Verbandsschuld oder Zurechnungsmodell?" (Culpability of the Corporation or Imputation Model?) in Martin Henssler, Elisa Hoven, Michael Kubiciel *et al.* (eds.), *Grundfragen eines modernen Verbandsstrafrechts* (Baden-Baden, Germany: Nomos, 2017) 71.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

inherent in criminal punishment can adequately be addressed to a legal person, an entity that has no conscience and cannot feel guilt.[33] An alternative basis for CCR could be a strictly functional approach to criminal law which links the responsibility of corporations to the empirical and/or normative expectation that they abide by the legal norms applying to their scope of activities.[34]

There exists an insoluble conflict between the pragmatic and political interest in nudging corporations toward legal compliance and the theoretical problems of extending the criminal law beyond natural persons. It is thus ultimately a policy question whether a state chooses to limit the liability of corporations for faults of their employees to tort law, extends it to criminal law, or places it somewhere in between,[35] as has been done in Germany.[36] In what follows, I assume that the criminal law version of CCR has been chosen. In that case, the further policy question arises as to whether CCR should include criminal responsibility for harm caused by AI devices used by the corporation.

III.B Legitimacy of CCR for Robots

As we have seen, retroactively identifying the fault of an individual human actor can be as difficult when an AI device was used as when some unknown employee of a corporation may have made a mistake.[37] The problem of allocating responsibility for robot action is further exacerbated by the black box element in self-teaching robots used on behalf of a corporation.[38]

[33] On these and other problematic aspects of CCR, see Thomas Weigend, "Societas delinquere non potest? A German Perspective" (2008) 6:5 *Journal of International Criminal Justice* 927. For ways of dealing with corporate misconduct outside the criminal law, see Charlotte Schmitt-Leonardy, *Unternehmenskriminalität ohne Strafrecht?* (Corporate Crime without Criminal Law?) (Heidelberg, Germany: C. F. Müller Verlag, 2013).

[34] As to that approach, see notes 12–18 above.

[35] See the strong argument in favor of "a softer version of the State's powers to prohibit and punish" in "AI Agents", note 22 above, at 696. The author plausibly warns that an overextension of criminal sanctions might "weaken our perception of what criminal law is and what it has the power to do."

[36] German law presently permits the imposition of administrative fines on corporations if their leading managers committed criminal offenses or culpably failed to prevent such offenses committed by employees; see *Gesetz über Ordnungswidrigkeiten* (Law on Administrative Infractions), of February 19, 1987, Germany, Bundesgesetzblatt 1987 I, 602, §§ 30, 130.

[37] See text at note 19 above.

[38] If the law treats robots like humans, CCR could be applied directly to robots' malfeasance. See e.g., the Michigan statute discussed by Clint W. Westbrook, "The Google Made Me Do It. The Complexity of Criminal Liability in the Age of Autonomous Vehicles"

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

It could be argued that the responsibility gap can be closed by treating the robot as a mere device employed by a human handler, which would turn the issue of a robot's harmful action into a regular instance of corporate liability. But even assuming that the doctrine of *respondeat superior* provides a sufficient basis for holding a corporation liable for faults of its employees, extending that doctrine to AI devices employed by humans would raise additional doubts about a corporation's responsibility. It may neither be known how the robot's harmful action came about nor whether there was a human at fault,[39] nor whether the company could have avoided the employee's potential malfeasance.[40] It is therefore unlikely that many cases of harm caused by an AI device could be traced back to recklessness or criminal negligence on the part of a human employee for whom the corporation can be made responsible.

Effectively bridging the responsibility gap would therefore require the more radical step of treating a company's robots like its employees, with the consequence of linking CCR directly to the robot's malfeasance. This step could set into motion CCR's beneficial compliance mechanism: if the robot's fault is transferred by law to the company that employs it, that company will have a strong incentive to design, program, and constantly monitor its robots to make sure that they function properly.

How would a corporation's direct responsibility for actions of its robots square with the general theories on CCR?[41] The alter ego-type liability model based on a transfer of the responsibility of employees to the corporation is not well suited to accommodating activities of robots because their actions lack the quality of blameworthy human decision-making.[42] Transfer of liability would work only if the mere existence of harmful

(2017) 2017:1 *Michigan State Law Review* 97 ["Google Made Me Do It"]. Michigan Compiled Laws s. 257.665(5), introduced in 2016, declares that an automated driving system is the driver or operator of a vehicle "for purposes of determining conformance to any applicable traffic or motor vehicle laws." From that legal provision, the author concludes that "manufacturers should be held liable for AV-caused crimes where their products are shown to be culpable for certain criminal acts and harm caused thereby" ("Google Made Me Do It," at 126), i.e., if a failure in hardware or software caused the infraction (ibid. at 133).

[39] "Criminal Liability of AI", note 22 above, at 62–63 correctly notes that strict liability for any malfeasance of a robot would place too heavy a burden on its individual programmers, designers, and distributors, eventually hampering the development of new technology.

[40] The cause of the harm could also lie in the robot's self-programming. As pointed out in "Algorithms Acting Badly", note 5 above, at 819–820, humans are increasingly absent from the process of writing code, with algorithms themselves writing most of the code for sophisticated programs.

[41] See text at notes 24–25 above.

[42] See "Impact of AI", note 19 above, at 148–149 and 153.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

activity on the part of an employee or robot would be sufficient to trigger CCR, i.e., in an absolute liability model. Such a model would address the difficulties raised by corporations using robots in situations where the robot's behavior is unpredictable; however, it is difficult to reconcile absolute liability with European concepts of criminal justice. A more promising approach to justifying CCR for robots relates to the corporation's overall spirit of lawlessness and/or its inherently defective organization as grounds for holding it responsible.[43] It is this theory that might provide an explanation for the corporation's liability for the harmful acts of its robots; if a corporation uses AI devices, but fails to make sure that they operate properly, or uses a robot when it cannot predict that the robot will act safely, there is good reason to impose sanctions on the corporation for this deficiency in its internal organization. This is true even where such AI devices contain elements of self-teaching. Who but the corporation that employs them should be able to properly limit and supervise this self-teaching function?

In this context, an analogy has been discussed between a corporation's liability for robots and a parent's or animal owner's liability for harm caused by children or domestic animals.[44] Even though the reactions of a small child or a dog cannot be completely predicted, it is only fair to hold the parent or dog owner responsible for harm that could have been avoided by training and supervising the child or the animal so as to minimize the risks emanating from them.[45] Similar considerations suggest a corporation's liability for its robots, at least where it can be shown that the robot had a recognizable propensity to cause harm. By imposing penalties on corporations in such cases, the state can effectively induce companies to program, train, and supervise AI devices so as to avoid harm.[46] Moreover, if there is insufficient liability for harm by robots, business firms might be tempted to escape traditional CCR by replacing human employees by robots.[47]

[43] See Kurt Schmoller, "'Verbandsschuld' als funktionsanaloges Gegenstück zur Schuld des Individualstrafrechts" ('Corporate Culpability' as a Functional Analogue to Culpability in Criminal Law for Individual Persons) in Marianne Johanna Lehmkuhl & Wolfgang Wohlers (eds.), *Unternehmensstrafrecht* (Basel, Switzerland: Helbing Lichtenhahn Verlag, 2020) 67.

[44] "AI Entities", note 6 above, at 1064–1066. Liability would normally be in tort law, but could also extend to criminal law, e.g., where an unsupervised dog bites a person.

[45] *Accord*, "Algorithms Acting Badly", note 5 above, at 809, 816, and 829 (claiming that "algorithmic action is corporate action"); "Criminal Liability of AI", note 22 above, at 71–72; "AI Entities", note 6 above, at 1067 and 1071 (arguing for treating robots as "agents").

[46] "Algorithms Acting Badly", note 5 above, at 831.

[47] Ibid. at 811.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

III.C Regulating and Limiting Robot CCR

Before embracing an extension of CCR from employees to robots, however, a counterargument needs to be considered. The increased deployment of AI devices is by and large a beneficial development, saving not only cost, but also human labor in areas where such labor is not necessarily satisfying for the worker, as in conveyor-belt mechanical manufacturing. Robots do have inherent risks, but commercial interests will provide strong incentives for their companies to control these risks. Adding criminal responsibility might produce an over-reaction, inhibiting the use and further development of AI devices and thus stifling progress. An alternative to CCR for robot malfunction may be for society to accept certain risks associated with the widespread use of AI devices and to restrict liability to providing compensation for harm through insurance.[48] These considerations do not necessarily preclude the introduction of a special regime of corporate liability for robots, but they counsel restraint. Strict criminal liability for robotic faults would have a chilling effect on the development of robotic solutions and therefore does not recommend itself as an adequate solution.

Legislatures should therefore limit CCR for robots to instances where human agents of the corporation were at least negligent with regard to designing, programming, and controlling robots.[49] Only if that condition is fulfilled can it be said that the corporation deserves to be punished because it failed to organize its operation so as to minimize the risk of harm to others. Potential control over the robot by a human agent of the corporation is thus a necessary condition for the corporation's criminal liability. Mihailis E. Diamantis plausibly explains that "control" in the context of algorithms means "the power to design the algorithm in the first place, the power to pull the plug on the algorithm, the power to modify it, and the power to override the algorithm's decisions."[50] But holding

[48] Cf. "AI Agents", note 22 above, at 694: "Not everything can be foreseen, prevented, or contained, and in everyday life there are several instances where no one is to blame – much more be held criminally liable – for an undesirable outcome … Not everything can or should be regulated under criminal law."

[49] Cf. "Algorithms Acting Badly", note 5 above, at 836; Dominik Schmidt & Christian Schäfer, "Es ist schuld?! – Strafrechtliche Verantwortlichkeit beim Einsatz autonomer Systeme im Rahmen unternehmerischer Tätigkeiten" (It's Its Fault?! – Criminal Responsibility in Connection with Employing Autonomous Systems in the Context of Entrepreneurial Activities) (2021) 10:11 *Neue Zeitschrift für Wirtschaftsstrafrecht* 413 at 420; "AI Agents", note 22 above, at 693.

[50] "Algorithms Acting Badly", note 5 above, at 835.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

every company that has any of these types of control liable for any harm that the robot causes, Diamantis continues, would draw the net wider than "sound policy or fairness would dictate."[51] He therefore suggests limiting liability for algorithms to companies which not only control a robot, but also benefit from its activities.[52] The combination of these factors is in fact perfectly in line with the requirements of traditional CCR, where liability presupposes that the corporation had a duty to supervise the employee who committed the relevant fault and that the employee's activity or culpable passivity was meant to benefit the corporation.

This approach appropriately limits CCR to corporations that benefit from the employment of AI devices. Even so, liability should not be strict in the sense that a corporation is subject to punishment whenever any of its robots causes harm and no human actor responsible for its malfunction can be identified.[53] In line with the model of CCR that is based on a dysfunctional organization of the corporation, criminal liability should require a fault on the part of the corporation that has a bearing on the robot's harmful activity.[54] This corporate fault can consist, e.g., in a lack of proper training or oversight of the robot, or in an unmonitored self-teaching process of the AI device.[55] There should in any event be proof that the corporation was at least negligent concerning its obligation to do everything in its power to prevent robots that work for its benefit from causing harm to others. In other words, CCR for robots is proper only where it can be shown that the corporation could, with proper diligence, have avoided the harm. This model of liability could be adopted even in jurisdictions that require some fault on the part of managers for CCR, because the task of properly training and supervising robots is so important that it should be organized on the management level.

Corporate responsibility for harm caused by robots differs from CCR for activities of humans and therefore should be regulated separately by statute. The law needs to determine under what conditions a corporation is to be held responsible for robot malfeasance. The primary issue that needs to be addressed is the necessary link between a corporation and

[51] Ibid. at 836.
[52] Ibid. at 844; "Criminal Liability of AI", note 22 above, at 69 also emphasizes the importance of the "benefit" element.
[53] *Accord*, "Criminal Liability of AI", note 22 above, at 693.
[54] For a similar concept in CCR, see *Strafgesetzbuch* (Swiss Criminal Code), SR 311.0 (as amended January 23, 2023), Art. 102, para. 2.
[55] For an overview of potential fault of human beings in connection with robots, see Chapter 1 in this volume.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press

an AI device. Taking an automated car as an example, there are several candidates for potential liability for its harmful operation: the firm that designed the car, the manufacturing company, the programmer of the software, the seller, and the owner of the car, if that is a corporation. If it can be proved that the malfunctioning of the car was caused by an agent of one of these companies, e.g., the programmer was reckless in installing defective software, that company will be liable under the normal CCR rules of the relevant jurisdiction. Special "Robot CCR" will come into play only if the car's aberration cannot be traced to a particular human source, for example, if the reason for the malfunction remains inexplicable even to experts, if there was a concurrence of several causes, or if the harmful event resulted from the car's unforeseeable defective self-teaching. In any of these instances, it must be determined which of the corporate entities identified above should be held responsible.

IV Conclusion

We have found that robots can at present not be subject to criminal punishment and cannot trigger criminal liability of corporations under traditional rules of CCR for human agents. Even if the reach of the criminal law is extended beyond natural persons to corporations, the differences between corporations and robots are so great that a legal analogy between them cannot be drawn. But it is in principle possible to extend the scope of corporate responsibility, including criminal responsibility if recognized in the relevant jurisdiction, to harm caused by AI devices controlled by corporations and operating for their benefit. Given the general social utility of using robots, however, corporate liability for harm caused by them should not be unlimited, but should at least require an element of negligence in programming, testing, or supervising the robot.

https://doi.org/10.1017/9781009431453.007 Published online by Cambridge University Press