Lernende Roboter und Fahrlässigkeitsdelikt

Tianyu Yuan*

Dieser Beitrag befasst sich mit der für die kognitive Robotik besonders bedeutsamen Methode des Maschinellen Lernens und analysiert dafür erforderliche menschliche Entscheidungen vor dem Hintergrund der Fahrlässigkeitsdelikte. Nach einführenden Bemerkungen zu grundlegenden Begrifflichkeiten (Roboter, Autonomie, Künstliche Intelligenz) erfolgt zunächst eine Beschreibung des maschinellen Lernprozesses unter Hervorhebung der vielfältigen damit verbundenen Programmierentscheidungen. Anschließend wird im Lichte des Fahrlässigkeitsdelikts untersucht, welche denkbaren Programmierfehler (im weiteren Sinne) in diesem Prozess auftreten können, und dargelegt, wie das bereits bestehende Recht Programmierer zu einem sorgfaltspflichtgemäßen Verhalten motiviert. Wenngleich autonome Agenten häufig als „Blackbox" bezeichnet werden und ihre Aktionen bisweilen unvorhersehbar erscheinen, darf der Mensch nicht vorschnell aus der (strafrechtlichen) Verantwortung genommen und Taterfolge nur dem Roboter „zugerechnet" werden. Der Beitrag schließt mit Gedanken zur Regulierung algorithmischer Entscheidungsprozesse unter Verwendung Maschinellen Lernens und hebt hervor, dass Algorithmen nicht per se, sondern nur im jeweiligen Anwendungszusammenhang zum Gegenstand straf- und ordnungsrechtlicher Regulierung werden sollten.

A. Einleitung

"Then you don't remember a world without robots. There was a time when humanity faced the universe alone and without a friend. Now he has creatures to help

* Tianyu Yuan ist akademischer Mitarbeiter am Institut für deutsches, europäisches und internationales Strafrecht und Strafprozessrecht (Lehrstuhl Prof. Dr. Jan C. Schuhr) an der Ruprecht-Karls-Universität in Heidelberg.

him; stronger creatures than himself, more faithful, more useful, and absolutely devoted to him. Mankind is no longer alone."[1]

Knapp 70 Jahre nachdem *Isaac Asimov* in „I, Robot" seinen Roboterprotagonisten diese Feststellung äußern ließ, scheinen wir in eine Zeit zu kommen, in der viele Menschen sich nur noch schwer werden vorstellen können, wie ein Leben ganz ohne Roboter und autonome Agenten funktionierte. Während Roboter in der industriellen Fertigung bereits seit den 1950ern im Einsatz sind,[2] werden sie seit einigen Jahren zunehmend Teil des menschlichen Alltags. Dies beginnt bei täglichen Trivialitäten wie Staubsaugen, Putzen oder Rasenmähen,[3] erreicht aber auch Bereiche komplexerer menschlicher Bedürfnisse.[4] Roboter haben längst ihre Industriekäfige verlassen[5] und reichen uns im täglichen Leben ihre helfende Hand.

Diese neue Intensität der Mensch-Maschine-Interaktion birgt neue Risiken, die an dieser Stelle im Lichte der Fahrlässigkeitsstrafe untersucht werden sollen. Dabei geht es nicht um die Analyse und Begründung einer maschinellen Rechtspersönlichkeit[6] oder der Bestrafung von Robotern,[7] sondern ganz herkömmlich um die juristische Beurteilung eines auf dem Dogma der Willensfreiheit[8] beruhenden menschlichen Verhaltens. Es soll versucht werden, die Implementierung eines maschinellen Lernprozesses als wichtigen Teilbereich der technischen Arbeitsschritte, mit welcher der Mensch die Maschine formt, in einer für die Subsumtion zugängli-

1 *I. Asimov*, I, Robot, New York 1950, S. 3.

2 Die erste Patentanmeldung für einen Industrieroboter erfolgte 1954 durch *George Devol* (US Patent Nummer 2988237A), der 1956 auch das weltweit erste Robotikunternehmen gründete; einen Überblick zur Geschichte der Robotik liefert *M. Haun*, Handbuch Robotik, Berlin 2013, S. 4 f.

3 Nach Schätzungen der International Federation of Robotics wurden 2017 knapp 6.1 Millionen Haushaltsroboter verkauft, IFR Executive Summary World Robotics 2018 Service Robots, S. 13; https://ifr.org/downloads/press2018/Executive_Summary_WR_Service_Robots_2018.pdf, zuletzt abgerufen am 15.2.2019.

4 Ein prominentes Beispiel ist „Paro" die Roboter Robbe, welche bereits seit 2004 vertrieben wird und zu therapeutischen Zwecken, insbesondere in der Altenpflege, zum Einsatz kommt; vgl. *T. Shibata/K. Wada*, Robot Therapy: A New Approach for Mental Healthcare of the Elderly, Gerontology 2011, S. 378 ff.

5 Dies betrifft auch Industrieroboter selbst, welche nun auch als „Cobot" (Abkürzung von „collaborative robot") im Produktionsprozess mit Menschen ohne physikalische Schutzeinrichtungen kollaborieren.

6 Vgl. z.B. *J. Kersten*, Die maschinelle Person – Neue Regeln für den Maschinenpark?, in: A. Manzeschke / F. Karsch (Hrsg.), Roboter, Computer, Hybride, Baden-Baden 2016, S. 89 (94 ff.); *A. Matthias*, Automaten als Träger von Rechten, Berlin 2008, S. 235 ff.; *G. Seher*, Intelligente Agenten als „Personen" im Strafrecht?, in: S. Gless/K. Seelmann (Hrsg.), Intelligente Agenten und das Recht, Baden-Baden 2016, S. 45 ff. jeweils mwN.

7 Vgl. z.B. *S. Gless/T. Weigend*, Intelligente Agenten und das Strafrecht, ZStW 2014, S. 561 (566 ff.); *M. Simmler/N. Markwalder*, Roboter in der Verantwortung?, ZStW 2017, S. 20 (32 ff.); *S. Ziemann*, Wesen, Wesen, seid's gewesen? Zur Diskussion über ein Strafrecht für Maschinen, in: E. Hilgendorf/J. Günther (Hrsg.), Robotik und Gesetzgebung, Baden-Baden 2013, S. 183 ff.

8 Mit kritischer Analyse, *C. Jäger*, Willensfreiheit, Kausalität und Determination – Stirbt das moderne Schuldstrafrecht durch die moderne Gehirnforschung?, GA 2013, S. 3 (6 ff.); in vergleichender Reflexion mit dem mathematischen Auswahlaxiom, *J. Schuhr*, Willensfreiheit, Roboter, Auswahlaxiom, in: S. Beck (Hrsg.), Jenseits vom Mensch und Maschine, Baden-Baden 2012,S. 8 ff.

chen Granularität darzustellen, um zu verstehen, wo bestehendes Strafrecht bereits heute regulierend wirken kann, und um einen Ausgangspunkt für strafrechtliche Überlegungen de lege ferenda zu schaffen. In strafrechtlicher Hinsicht ist die Fahrlässigkeitsstrafe von besonderem Interesse, da vorsätzliche Taten unter Verwendung von Robotern als Werkzeug sich materiell-rechtlich ohne besondere Herausforderungen handhaben lassen und zu vermuten steht, dass in der Rechtspraxis der Fahrlässigkeitsvorwurf bei Programmierfehlern eine größere Rolle spielen wird.[9] Bevor allerdings die juristische Analyse erfolgen kann, bedarf es einiger Ausführungen zum Sprachgebrauch in Bezug auf den Gegenstandsbereich und einer Beschreibung des Gegenstandsbereichs des Maschinellen Lernens selbst.

1. Roboter

Regelmäßig werden mit „Roboter" Maschinen bezeichnet, die über Sensoren ihre physikalische Umgebung wahrnehmen, diese Daten mittels Prozessoren verarbeiten und über Aktuatoren auf ihre Umgebung physisch einwirken.[10] Wie bei allen übergeordneten Begriffen, die sich auf eine dynamische tatsächliche Entwicklung beziehen, unterliegt auch der Roboterbegriff semantischen Veränderungen.[11] Für die juristische Handhabung würden diese insoweit interessieren, wie entweder das Gesetz tatbestandlich an den Begriff des Roboters anknüpft, was allerdings in der bundesdeutschen Gesetzgebung bislang nur in Anlagen zu Gesetzen und Rechtsverordnungen geschehen ist,[12] oder Roboter unter übergeordnete Begriffe zu subsumieren sind, wie z.B. der „Maschine" im Anwendungsbereich der Maschinenrichtlinie (RL 95/16/EG vom 17. Mai 2006), die in Deutschland durch das Produktsicherheitsgesetz und die dazugehörige Maschinenverordnung umgesetzt wurde.[13]

Die International Federation of Robotics knüpft an ISO 8373:2012 an und unterteilt Roboter je nach Einsatzgebiet in oder außerhalb der Automatisierungstechnik

9 Ähnlich *Gless/Weigend*, Intelligente Agenten (Fn. 7), S. 579 f.

10 Anthropomorph zugespitzt: "we define a robot as a machine that senses, thinks and acts", *G. Bekey*, Autonomous Robots, Cambridge, Mass.: The MIT Press 2005, S. 2.

11 So wird der ISO-Standard 8373:2012 aus dem Jahre 2012 für Roboter, die sowohl in industriellen als auch in nicht-industriellen Umgebungen zum Einsatz kommen, derzeit überarbeitet und künftig durch ISO/CD 8373 ersetzt.

12 Z.B. in Anlage 1 Anlage AL zur Außenwirtschaftsverordnung, die auch eine Begriffsbestimmung des Roboters enthält. Danach ist ein Roboter „ein Handhabungssystem, das bahn- oder punktgesteuert sein kann, Sensoren benutzen kann und alle folgenden Eigenschaften aufweist: a) multifunktional, b) fähig, Material, Teile, Werkzeuge oder Spezialvorrichtungen durch veränderliche Bewegungen im dreidimensionalen Raum zu positionieren oder auszurichten, c) mit drei oder mehr Regel- oder Stellantrieben, die Schrittmotoren einschließen können, und d) mit „anwenderzugänglicher Programmierbarkeit" durch Eingabe-/Wiedergabe-Verfahren (teach/playback) oder durch einen Elektronenrechner, der auch eine speicherprogrammierbare Steuerung sein kann, d. h. ohne mechanischen Eingriff.".

13 Inwieweit „Roboter" Gegenstand der Rechtsprechung waren, s. *J. Schuhr*, Recht, Technik, Roboter, RT 2015, S. 225 (225 f.).

in Industrieroboter und Serviceroboter.[14] Letztere sind solche, die „nützliche Aufgaben für Menschen, die Gesellschaft oder Einrichtungen verrichte[n], mit Ausnahme von Anwendungen in der Automatisierungstechnik."[15] Damit decken Serviceroboter einen weiten Bereich ab, der sowohl Einsätze im Privaten als auch gewerblichen oder beruflichen Kontext erfasst.[16] Ferner können Roboter auch in Beziehung zur Komplexität ihrer Aktionsumgebung unterteilt werden.[17] Eine solche Einteilung spiegelt gleichzeitig die Entwicklungsgeschichte der Robotik wieder. Angefangen bei nicht-mobilen Industrierobotern, die für eine spezifische Aufgabe in einer eng definierten Umgebung tätig waren, kommen neuere Industrie- und Serviceroboter zunehmend mobil und in weniger klar ex ante definierten Umgebungen zum Einsatz und interagieren mit Menschen und anderen Robotern. Roboter, die in dynamischen und uneindeutigen Situationen agieren und deren Einwirkung auf die Umgebung von Unsicherheit geprägt sind, werden als kognitive Roboter bezeichnet.[18] Gerade in der kognitiven Robotik werden Erkenntnisse aus dem Forschungsbereich der Künstlichen Intelligenz (KI) implementiert. Damit wechselt der Fokus zunehmend von der Hardware zur Software: Roboter sind verkörperte KI. Aufgrund dieser Schwerpunktverschiebung soll deshalb im Folgenden auch schlicht von „Agenten" die Rede sein, womit sowohl Roboter als auch reine Softwareagenten gemeint sind.

2. Autonomie

Im Zusammenhang mit kognitiven Robotern fällt regelmäßig auch der Begriff der Autonomie. So ist häufig von „autonomen Agenten" die Rede und selbstfahrende Fahrzeuge werden bisweilen als „autonom" bezeichnet. Dieser bedeutungsmächtige Begriff, in dem *Kant* den „Grund der Würde der menschlichen und jeder vernünftigen Natur"[19] sah, birgt im technischen Kontext allerdings eine gewisse Ge-

14 IFR Topics and Definitions, https://ifr.org/#topics, zuletzt abgerufen am 15.2.2019; dazu auch *M. Müller*, Roboter und Recht, AJP 2014, S. 595 (596).

15 Service robot nach ISO 8373:2012: „A robot that performs useful tasks for humans or equipment excluding industrial automation application.", deutsche Übersetzung s. *Müller*, Roboter (Fn. 14), S. 596 f.

16 Eine instruktive Aufzählung findet sich in World Robotics 2016, Service Robots, S. 11 f., Zugriff über https://ifr.org/service-robots, zuletzt abgerufen am 15.2.2019. Ob eine solche weitere Einteilung der Service Roboter nicht mehr nach Einsatzgebiet, sondern nach privatem oder gewerblichem bzw. beruflichem Verwendungszweck sinnvoll erscheint, darf bezweifelt werden. Jedenfalls erfüllt sie die definitorische Abgrenzungsfunktion nur unzureichend, weil derselbe Roboter wie z.B. ein autonomes Fahrzeug, das sowohl privat als auch beruflich genutzt wird, je nach Zwecksetzung des Nutzers die Kategorie wechseln kann.

17 *Haun*, Robotik (Fn. 2), S. 29.

18 *Haun*, Robotik (Fn. 2), S. 29; *J. Hertzberg*, Kognitive Robotik, in: A. Stephan/S. Walter (Hrsg.), Handbuch Kognitionswissenschaft, Stuttgart 2013, S. 47 (47).

19 *I. Kant*, Grundlegung zur Metaphysik der Sitten, 1785 aus: Akademieausgabe von Immanuel Kants Gesammelten Werken, Band IV, Berlin 1968, S. 436, Z. 6 f.

fahr der Irreführung[20] und könnte dazu motivieren, strafrechtlich relevante Erfolge nur der Maschine „zuzurechnen“,[21] während der Mensch aus der Verantwortung genommen wird.[22] Wenn Ingenieure oder Programmierer von autonomen Robotern oder autonomen Agenten sprechen, bilden sie sich keinesfalls ein, dem Gegenstand ihrer Arbeit Menschlichkeit oder einen freien Willen verliehen zu haben. Sondern mit Autonomie wird im Zusammenhang mit Robotern die Fähigkeit bezeichnet, in unbekannten – also nicht explizit a priori im Programm definierten – Umgebungen im Sinne der festgelegten Zielsetzung zu agieren, indem durch Sensoren die Umgebung nach und nach erfasst wird und die Aktionen auf Grundlage des neuen Umgebungswissens angepasst werden.[23] Zur deutlicheren Differenzierung zwischen menschlicher und technischer Autonomie wird deshalb vorgeschlagen, die unterschiedlichen Autonomietypen besonders zu kennzeichnen oder in ein Stufenverhältnis zu setzen.[24] In Ansehung dieser Unterschiede soll deshalb im Zusammenhang mit durch Software motivierten Roboterbewegungen auch nur von „Aktionen“ bzw. „agieren“ die Rede sein und der Begriff des „Verhaltens“ gemieden werden, da letzterer einen Bezug zur strafrechtlichen Handlungslehre aufweist, die sich nur auf Menschen bezieht.[25]

3. Künstliche Intelligenz

Im Zusammenhang mit der „Autonomie“ fällt häufig auch der Begriff der „Künstlichen Intelligenz“, um die besondere Eigenständigkeit von Robotern hervorzuheben, was ebenfalls eine gewisse „Eigenverantwortlichkeit“ ihrer Aktionen suggeriert. Bei einer naiven Annährung an den Begriff der KI könnte man vertreten, „künstlich“ bezeichne im Sinne eines Artefakts einfach etwas, das von Menschen geschaffen wurde und „intelligent“ bedeute die Fähigkeit, besonders effizient Probleme lösen zu können, sodass man es bei KIs mit Artefakten, die besonders effizient Probleme lösen können, zu tun hat. Weil aber „Intelligenz“ einerseits häufig als

20 Ausführlich zur unterschiedlichen Bedeutung der Autonomie im geisteswissenschaftlichen und technischen Kontext, s. *M. Müller*, Von vermenschlichten Maschinen und maschinisierten Menschen, in: S. Brändli/ R. Harasgama/R. Schister/A. Tamò (Hrsg.), Mensch und Maschine – Symbiose oder Parasitismus?, Bern 2014, S. 125 (130 ff.).

21 Über den Sinn einer Übertragung des Zurechnungsbegriffs auf Roboter, *J. Schuhr*, Neudefinition tradierter Begriffe (Pseudo-Zurechnungen an Roboter), in: E. Hilgendorf (Hrsg.), Robotik im Kontext von Recht und Moral, S. 13 ff.

22 Im Zusammenhang mit der Zurechnung im Rahmen von Fahrlässigkeitsdelikten, s. Teil C., 4. unten.

23 Vgl. *S. Russel/P. Norvig*, Artificial Intelligence, Upper Saddle River, New Jersey: Prentice Hall 2010, S. 39; nach *Hilgendorf* soll „autonom“ im technischen Sinne etwas weiter gefasst das Agieren “unabhängig von menschlichen Eingaben im Einzelfall“ bedeuten, *E. Hilgendorf*, Können Roboter schuldhaft handeln?, in: S. Beck (Hrsg.), Jenseits von Mensch und Maschine, Baden-Baden 2012, S. 119 (120, Fn. 2).

24 *Müller*, vermenschlichte Maschinen (Fn. 20), S. 139 f. mwN.

25 Zur strafrechtlichen Handlungslehre anstelle vieler: *T. Fischer*, Strafgesetzbuch, 66. Aufl., München 2019, Vor § 13 Rn. 3 ff.; *C. Roxin*, Strafrechtliche Grundlagenprobleme, Berlin 1973, S. 72 ff. jeweils mwN.

menschliches Attribut gesehen und andererseits mit tugendhafter Rationalität assoziiert wird, setzen Definitionsversuche der „Künstlichen Intelligenz" auch dort an und bezeichnen Agenten als „künstlich intelligent", die jeweils entweder „menschenähnlich denken", „menschenähnlich agieren", „rational denken" oder „rational agieren" können.[26]

Mit der KI ist auch der Gegenstandsbereich dieser Untersuchung erreicht, welche die Frage zu beantworten versucht, inwieweit Programmierer im Zusammenhang mit Roboteraktionen, die auf Maschinellem Lernen beruhen, sich wegen Fahrlässigkeitsdelikten zu verantworten haben. Die Fähigkeit von Agenten, zu einem gewissen Grade selbständig zu lernen, stellt eine wesentliche Eigenschaft moderner KI-Systeme dar. Man kann durchaus behaupten, dass die heutige Wissensrepräsentation als wichtiger Teil der KI-Forschung und -Praxis durch das Maschinelle Lernen geprägt ist, welcher durch die zunehmende Verfügbarkeit großer Datenmengen (Big Data) begünstigt bzw. überhaupt erst ermöglicht wird.[27] Zuvor, in den 1970er und 80er Jahren, setzte die Praxis intensiv auf sog. Expertensysteme, welche in einem in der Regel eng definierten Problemumfeld von menschlichen Domänenexperten mit explizitem Wissen ausgestattet wurden und durch im Voraus klar definierten Wenn-Dann-Beziehungen Schlussfolgerungen ziehen, entsprechend agieren und damit Probleme lösen konnten.[28] Expertensysteme spielen auch heute noch eine wichtige Rolle und werden z.B. in der Energieversorgung eingesetzt.[29] In der juristischen Praxis scheinen professionelle Expertensysteme allerdings erst seit wenigen Jahren im Zusammenhang mit dem wachsenden Interesse an „Legal Tech" Einzug zu finden.[30]

Der Fokuswechsel von Expertensystemen zu maschinell lernenden Systemen ist nicht nur durch die Verfügbarkeit großer Datenmengen, sondern auch durch die zunehmende Komplexität der Problemumgebung bedingt. Außerhalb industrieller Fertigungsstraßen, in Zusammenarbeit mit Menschen und in Alltagssituationen,

26 „Denken" bezeichnet an dieser Stelle natürlich schlicht die prozessorgestützte Datenverarbeitung und Menschenähnlichkeit und Rationalität sind ihrerseits stark wertungsoffene Ausdrücke; zu den verschiedenen Definitionsansätzen s. *Russel/Norvig*, AI (Fn. 23), S. 2 mwN. Bisweilen wird als fundamentales Problem des Begriffs der KI auch angemerkt, dass ohnehin niemand wisse, was „Intelligenz" bedeute, vgl. *S. Legg/M. Hutter*, Universal Intelligence: A Definition of Machine Intelligence, Mind and Machines 2007, S. 391 (391): „A fundamental problem in artificial intelligence is that nobody really knows what intelligence is.".

27 Zum Zusammenhang zwischen Data Mining und Machine Learning, S. *I. Witten/E. Frank*, Data Mining, Amsterdam 2005, S. 4 ff.

28 Zum geschichtlichen Überblick s. *Russel/Norvig*, AI (Fn. 23), S. 22 ff.; im juristischen Kontext bereits *R. Susskind*, Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning, Modern Law Review 1986, S. 168 ff.

29 *Z. Styczynski/K. Rudion/A. Naumann*, Einführung in Expertensysteme, Berlin 2017, S. 14 ff.

30 Praxistools sind z.B. aus dem anglo-amerikanischen Bereich Neota Logic, das seit 2010 entwickelt wird, oder aus dem deutschsprachigen Raum BRYTER, das erst 2018 im Rechtsmarkt bekannter wurde.

können Serviceroboter nicht mehr durch explizite Programmieranweisungen im Vorfeld auf alle erdenklichen Situationen vorbereitet werden – man denke nur an die Vielfalt von Verkehrssituationen, die ein autonomes Fahrzeug bewältigen können muss. Weil konkrete Situationen, in denen sich Roboter wiederfinden können, sehr komplex und dadurch häufig nur bedingt vorhersehbar sind, wird auch das Aufstellen expliziter Aktionsregeln deutlich erschwert. Das Vorstellungsvermögen und die prädiktive Kraft der Menschen stoßen an ihre Grenzen, nicht aber ihre kreativen Problemlösungsfähigkeiten: Im Umgang mit Komplexität und Unvorhersehbarkeit wurde im Maschinellen Lernen eine Lösung gefunden.[31]

B. Maschinelles Lernen

Wenn Maschinelles Lernen im Zusammenhang mit algorithmischen Entscheidungsverfahren (Algorithmic Decision-Making, ADM) im Spiel ist, wird häufig die Metapher der „Black Box" bemüht, um zum Ausdruck zu bringen, dass bei solchen Verfahren unter Umständen der Zusammenhang zwischen Eingabe- und Ausgabewerten nicht oder nur begrenzt erklärt werden kann.[32] Gerade weil das gesamte Forschungsfeld um KI, ADM und Maschinellem Lernen sehr komplex und dynamisch ist, kann im vorliegenden Beitrag nur eine überblicksartige Annäherung versucht werden. Zudem gilt es, der Versuchung zu widerstehen, beim Stichwort „Black Box" und in einem technischen bzw. mathematischen Kontext vorschnell den Menschen für etwaige Rechtsgutverletzungen aus der (strafrechtlichen) Verantwortung zu nehmen, weil etwa der Erfolg nicht hätte vorhergesehen werden können oder er den handelnden Menschen nicht zurechenbar wäre.

Auch wenn der Begriff der „Black Box" im juristischen Sprachgebrauch selten vorkommt, sind Juristen im Umgang mit Black Box-Situationen alles andere als unerfahren. Denn bei jedem Subsumtionsvorgang muss der Jurist entscheiden, inwieweit der zu untersuchende Zusammenhang zwischen Rechtsbegriff und Lebenssachverhalt zu durchleuchten ist und an welcher Stelle eine weitere Analyse des Sachverhalts im Dunkeln bleiben kann. Es geht um die Entscheidung, inwiefern ein Rechtsbegriff explikativ zu definieren ist und welche Merkmale der Definition in

31 Eine entfernt verwandte, aber Juristen vertraute Situation des Umgangs mit der begrenzten menschlichen Fähigkeit, Zukunft vorherzusehen, kann auch in der Gesetzgebung gesehen werden, bei der – ungeachtet einer üblichen Einteilung in „deskriptive" oder „normative" Merkmale – bewusst abstrakte Begriffe gewählt werden, um Regelungsbereiche zu umschreiben. Auch in diesem Fall wird darauf vertraut, dass das Rechtssystem im konkreten Fall Begriffe mit Leben füllen kann und mit Gesetzesänderungen interveniert, wenn die Anwendung im Einzelfall nicht in dem Sinne wirkt, wie der Gesetzgeber sie für richtig hält.

32 Vgl. z.B. *E. Schweighofer et al.*, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren, Berlin 2018, S. 30 f.; zu Ansätzen und Methoden Erklärbarkeit unter Einsatz von ADM zu erzielen, s. *B. Waltl/R. Vogl*, Explainable Artificial Intelligence – the New Frontier in Legal Informatics, in: E. Schweighofer/F. Kummer/A. Saarenpää/B. Schafer (Hrsg.), Tagungsband zur IRIS 2018, S. 113 ff.; *P. Adler et al.*, Auditing black-box models for indirect influence, Knowledge and Information Systems 2018, S. 95 ff.

weiteren Schritten jeweils zu explizieren sind. Aus dem Blickwinkel des Rechts betrachtet, muss ein Sachverhalt nur insoweit verstanden werden, wie die rechtlichen Voraussetzungen es erfordern.[33] Die gesamte weitere Komplexität, welche der Sachverhalt in tieferen Ebenen oder anderen Aspekten aufweist, darf eine Black Box bleiben. Damit darf der folgende Blick auf das Maschinelle Lernen nicht nur im Sinne eines Überblicks ausgerichtet sein, sondern muss zumindest jene Aspekte beleuchten, die bei der folgenden Beurteilung der Fahrlässigkeitsstrafbarkeit wieder aufgegriffen werden.

1. Aufgabe

Wie der Begriff des Maschinellen Lernens bereits vorgibt, besteht die Aufgabe darin, algorithmische Modelle zu entwickeln, die für ein bestimmtes Ziel trainiert werden können und die Fähigkeit besitzen, sukzessiv durch weitere Eingaben oder Beobachtungen die Zielerreichung zu verbessern.[34] Technischer gesprochen werden dem Algorithmus im Rahmen des Lernprozesses Beispieldaten (Samples) mit bestimmten Eigenschaften (Variablen, Attributen oder Features) präsentiert und dieser soll für eine neue, unbekannte Instanz als Datum Vorhersagen treffen (Prediction).[35] D.h. die neue Instanz liefert die Eingabedaten (Input) und der Algorithmus verbindet sie mit bestimmten Ausgabedaten (Output). Durch den Algorithmus erfolgt damit ein sog. Mapping von Input und Output, wobei der lernende Agent diese Zuordnung im Laufe der Zeit immer besser beherrschen soll.

Dabei bedarf die Verwendung des Begriffs der „Vorhersage" einer Präzisierung. Vorhersage meint nicht, dass ein bestimmter Input als Ereignis eintreten und damit der Fall sein wird. Sondern es geht um die Situation, dass ein bestimmtes Ereignis als Input bereits eingetreten ist. Die Vorhersage bezieht sich nur auf den Output und ggf. noch auf andere Modelle, die an diesen Output anknüpfen, um daraus Aussagen über künftige Ereignisse anderer Art abzuleiten.[36] Als Beispiel: Wenn der Input darin besteht, dass ein Mensch einen anderen vorsätzlich, rechtswidrig und schuldhaft getötet hat, würde als Output die „Vorhersage" getroffen werden können, dass dieser Mensch sich wegen Totschlags (§ 212 Abs. 1 StGB) strafbar ge-

33 Ein solcher Blickwinkel soll nur verdeutlichen, dass die Black Box im juristischen Arbeitsprozess natürlicherweise vorkommt. Zur Beschreibung des Subsumtionsvorgangs ist eine einseitige, auf den jeweiligen Rechtsbegriff begrenzte Sicht natürlich unvollständig, weil dieser mit dem zu subsumierenden Ereignis in besonderer Weise verbunden ist. Einprägsam hat *Engisch* den Subsumtionsvorgang als eine „ständige Wechselwirkung [und als] ein Hin- und Herwandern des Blickes zwischen Obersatz und Lebenssachverhalt" bezeichnet, *K. Engisch*, Logische Studien zur Gesetzesanwendung, 3. Aufl., Heidelberg 1963, S. 15.

34 Vgl. *Russel/Norvig*, AI (Fn. 23), S. 693; *Witten/Frank*, Data Mining (Fn. 27), S. 6.

35 Siehe scikit-learn user guide, Release 0.20.2 vom Dez. 2018, S. 127; abrufbar über: https://scikit-learn.org/stable/documentation.html. scikit-learn ist eine viel genutzte und umfangreiche Software-Bibliothek für das Maschinelle Lernen in der Programmiersprache Python.

36 Zur Prediction und juristischen Anwendungsbeispielen vgl. *K. Ashley*, Artificial Intelligence and Legal Analytics, Cambridge, UK: Cambridge Press 2017, S. 108 ff.

macht hat. Ein daran anknüpfendes Modell würde z.B. vorhersagen, dass die Staatsanwaltschaft ermitteln wird, wenn sie davon Kenntnis erlangt (§ 160 Abs. 1 StPO). D.h. die Vorhersage betrifft gerade nicht, dass die Umstände eines Totschlags eintreten werden.[37]

2. Typen

Regulär werden maschinelle Lernverfahren in überwachtes Lernen (Supervised Learning), unüberwachtes Lernen (Unsupervised Learning) und verstärkendes Lernen (Reinforcement Learning) eingeteilt, wobei Mischformen, wie das Semi-Supervised Learning, häufig vorkommen.[38] Die Unterschiede liegen insbesondere in der Art des Feedbacks, anhand dessen der Algorithmus trainiert wird, und wodurch er das festgelegte Ziel besser erreichen soll.[39] Weitere Unterschiede liegen aber auch darin, welche Eingabedaten verfügbar sind oder welche Form die Ausgabedaten haben sollen – es geht also einerseits um die verfügbaren Datenressourcen und andererseits um das zu lösende Problem an sich.

a) Supervised Learning

Beim überwachten Lernen erhält der Algorithmus einen Trainingsdatensatz, bei dem die gesuchten Output-Eigenschaften der Daten (Labels) bereits bekannt sind (gelabelte Daten). Anhand dieses Trainingsdatensatzes soll der Algorithmus eine Funktion finden, der die Input-Output-Relation der Daten möglichst präzise beschreibt und bei einem neuen Datum mit bekannten Input-Werten, den Output vorhersagt. [40] Dabei muss das Training mit dem Ausgangsdatensatz natürlich nicht abgeschlossen sein, sondern mit jedem weiteren gelabelten Datensatz kann der Algorithmus einen weiteren Trainingszyklus durchlaufen, wodurch er sich sukzessive verbessert – er lernt.

Das Supervised Learning befasst sich häufig mit Problemen, die durch eine Klassifikation oder Regression charakterisiert werden. Bei der Klassifikation gehört der Output einer oder mehreren Kategorien an, und der Algorithmus soll anhand der Input-Daten diese Zuordnung treffen. Klassische Anwendungsfälle wären z.B. die Zuordnung, ob eine E-Mail Spam oder kein Spam ist, ob ein bestimmter Pixelbereich einer elektronischen Bilddatei ein KfZ darstellt oder nicht, oder die Typisierung von Schmetterlingen anhand von Flügelspannweite, Farbe und/oder anderer Eigenschaften hinsichtlich bestimmter Schmetterlingsarten. D.h. der Output ist ent-

37 Das bedeutet natürlich nicht, dass man nicht anhand anderer Inputs Modelle entwerfen kann, welche die Prädiktion der zuvor genannten Inputs – den Umständen des Totschlags – zum Ziel haben soll. Inwieweit ein solches Unterfangen erfolgsversprechend ist, kann dahingestellt bleiben.

38 Dies liegt daran, dass für das Supervised Learning häufig nicht genügend annotierte Daten vorliegen bzw. eine umfangreiche Annotation zu aufwendig und damit zu teuer ist, vgl. *Russel/Norvig*, AI (Fn. 23), S. 695.

39 *Russel/Norvig*, AI (Fn. 23), S. 694 f.

40 *Russel/Norvig*, AI (Fn. 23), S. 695 f.; *M. Hoogendoorn/B. Funk*, Machine Learning for the Quantified Self, Cham (CH) 2018, S. 7.

weder binär (Spam/kein Spam; Auto/kein Auto) oder durch abzählbare Typen (die einzelnen Schmetterlingsarten, die im konkreten Fall von Interesse sind) strukturiert.[41] Regressionen befassen sich mit numerischen Output-Werten, wie z.B. im Falle der Vorhersage eines Immobilienpreises in Abhängigkeit von der Wohnfläche und/oder anderen Faktoren als Input-Werten.[42] Für den juristischen Anwendungsbereich werden damit Klassifikationsalgorithmen besonders relevant, weil rechtliche Entscheidungen sich mit Klassifikationsproblemen befassen: Es geht um die Frage, ob ein bestimmtes gesetzliches Merkmal – sei es auf Tatbestands- oder Rechtsfolgenseite – erfüllt ist oder nicht. In der Terminologie des maschinellen Lernens gesprochen: Subsumtion ist Klassifikation.

Um ein Supervised Learning-Beispiel im Zusammenhang mit Fahrlässigkeitsdelikten zu bilden: Wir gehen von einem Datensatz aus, das aus einzelnen Absätzen von Entscheidungsgründen besteht, welche ausschließlich aus Entscheidungen über Fahrlässigkeitsstraftaten stammen. Eine Klassifikationsaufgabe könnte nun in der „Vorhersage“[43] liegen, ob ein zu klassifizierendes Datum – d.h. ein einzelner Absatz – eine Definition, einen Streitstand, Kasuistik oder einfach etwas anderes enthält, wobei eine Mehrfachklassifikation zulässig ist. Im Falle des Supervised Learning bräuchte man nun eine möglichst große Zahl von Trainingsdaten, bei denen bereits feststeht, welche Labels sie tragen. Der Algorithmus soll nun daraus eine Abbildungsfunktion gewinnen, mit der eine neue Instanz, die nicht im Trainingsdatensatz enthalten ist, automatisch gelabelt wird. Daraus ließe sich z.B. ein automatisiertes Annotationstool für Entscheidungsgründe zur Fahrlässigkeitsstrafe programmieren.

b) Unsupervised Learning

Im Falle des Unsupervised Learning haben wir es dagegen mit einem Datensatz zu tun, bei dem die Input-Daten zwar bekannt sind, der Output aber nicht exakt feststeht. Die Aufgabe des Algorithmus besteht nun darin, selbständig Zusammenhänge zwischen den Daten, also die Features, zu finden. Ein Feedback in dem Sinne, dass dem Algorithmus „gesagt“ wird, ob er etwas „richtig“ oder „falsch“ zugeordnet hat, sind für diesen Typus algorithmischen Lernens nicht vorgesehen.[44]

Einer der wichtigsten Anwendungsfälle des Unsupervised Learning liegt im sog. Clustering: Durch Clustering-Algorithmen werden die Daten zu bestimmten Grup-

41 Vgl. *Russel/Norvig*, AI (Fn. 23), S. 696; begrifflich etwas irreführend kommt allerdings als mathematische Funktion für Klassifikationsaufgaben die sog. „logistische Regression“ zum Einsatz, deren Eingangswerte zwar numerisch sind, der davon abhängige Zielwert allerdings binär interpretiert werden kann.

42 Vgl. *Russel/Norvig*, AI (Fn. 23), S. 696.

43 Gerade hier zeigt sich, dass der Begriff der „Vorhersage“ etwas „überdimensioniert“ ist, weil es schlicht um die Vornahme einer Klassifikation geht.

44 *Russel/Norvig*, AI (Fn. 23), S. 694 f., 817 f.; *Hoogendoorn/Funk*, Machine Learning (Fn. 40), S. 7.

pen (Cluster) zusammengefasst, bei denen der Algorithmus Ähnlichkeiten festgestellt hat.[45] Clustering kann einerseits nicht-hierarchisch stattfinden. Die Aufgabe besteht dann darin, einen Datensatz in eine vordefinierte Anzahl von Gruppen zu „zerlegen", die sich nicht überschneiden.[46] Andererseits ist auch hierarchisches Clustering möglich. Dabei kann entweder zunächst ein Supercluster gebildet werden, das anschließend auf mehreren Ebenen in Subcluster unterteilt wird; oder man beginnt bei Clustern auf tiefster Ebene und fasst diese zu Clustern höherer Ebenen zusammen.[47] So kann ein nicht-hierarchischer Clustering-Algorithmus z.B. aus einem Datensatz von Verkehrsbildern Pixelgruppen als ähnlich gruppieren und dadurch Features bilden, denen der Mensch im Nachhinein die Namen „Auto", „Fußgänger", „Fahrradfahrer" oder „Baum" gibt. Ein hierarchischer Clustering-Algorithmus könnte aus einem Supercluster, dem die Bezeichnung „Auto" zugewiesen wird, weitere Subcluster bilden, die dann als „Scheinwerfer", „Seitenspiegel" oder „Windschutzscheibe" bezeichnet werden.[48]

Im Falle des oben eingeführten Beispiel-Datensatzes würde ein non-hierarchischer Clustering-Algorithmus prinzipiell Cluster bilden können, die jeweils einen inhaltlichen Bezug zu einzelnen Merkmalen des Fahrlässigkeitsdelikts, wie z.B. der Sorgfaltspflichtverletzung oder der Vorhersehbarkeit, aufweisen. Ein hierarchischer Clustering-Algorithmus würde zusätzlich beispielsweise die Sorgfaltspflichtverletzung und die Vorhersehbarkeit agglomerativ zu einem Supercluster zusammenfassen, der sich im Gesetz unter dem Begriff „Fahrlässigkeit" (im StGB z.B. bei §§ 222 und 229) wiederfindet.

c) Reinforcement Learning

Das verstärkende Lernen setzt darauf, dass ein Algorithmus auf Grundlage eines verstärkenden Feedbacks (Reinforcement) trainiert wird, das sowohl positiv (Reward) als auch negativ (Punishment) konstruiert werden kann.[49] Reinforcement Learning kommt insbesondere in komplexen Domänen zum Einsatz, die von Entscheidungsabläufen geprägt sind, bei denen nicht stets eindeutig gesagt werden kann, ob eine einzelne Aktion „richtig" oder „falsch" war. Dadurch eignen sich solche Szenarien nicht für das überwachte Lernen, weil den erforderlichen Trainingsdaten eine hohe Unsicherheit anhaften würde.[50] Das Belohnungsmodell ist von der jeweiligen Problemstellung abhängig: Besteht beispielsweise die Aufgabe

45 *Hoogendoorn/Funk*, Machine Learning (Fn. 40), S. 7.

46 *Hoogendoorn/Funk*, Machine Learning (Fn. 40), S. 82 f.

47 *Hoogendoorn/Funk*, Machine Learning (Fn. 40), S. 84 ff.

48 Zum Clustering von Bilddaten im Zusammenhang mit Fahrzeugen, z.B. *E. Ohn-Bar/M. Trivedi*, Learning to Detect Vehicles by Clustering Appearance Patterns, IEEE Transactions on Intelligent Transportation Systems 2015, S. 2511 ff.

49 *Russel/Norvig*, AI (Fn. 23), S. 695, 830.

50 *Russel/Norvig*, AI (Fn. 23), S. 830 f.

darin, ein Schachspiel zu gewinnen, würde das Modell an den jeweils geschlagenen Figuren Reinforcements anknüpfen; im Training eines autonomen Fahrzeugs, das lernen soll, die Fahrspur zu halten, könnte man z.B. an die zurückgelegte Strecke ein „Reward" und an jede Überschreitung der Fahrbahnbegrenzung ein „Punishment" anknüpfen.[51]

Tatsächlich wurde 2018 erstmalig demonstriert, wie ein autonomes Fahrzeug allein durch Reinforcement Learning innerhalb eines Tages das Spurhalten gelernt hat.[52] Was in technologischer Hinsicht sicherlich als Durchbruch qualifiziert werden kann, darf aber nicht zu der Vorstellung verleiten, dass Maschine und Algorithmus all dies allein bewerkstelligt haben. Auch hier korrespondiert Learning mit Training und beruht damit letztlich auf menschlichem Verhalten. Was der Lernalgorithmus als Belohnung oder Bestrafung ansieht, ist das Ergebnis einer menschlichen Entscheidung. Außerdem muss auch der Mensch definieren, welche Aspekte der Umgebung der Agent überhaupt wahrnehmen kann (State Space) und welche Aktionen möglich sind (Action Space). Im erwähnten Beispiel haben die Entwickler den State Space durch Kamerabild, Fahrzeuggeschwindigkeit und Lenkradposition bestimmt; der Action Space umfasste die Veränderung der Lenkradposition sowie die Erreichung einer bestimmten Sollgeschwindigkeit; und das Belohnungsmodell richtete sich nach der zurückgelegten Strecke bis ein menschlicher Trainer das Lenkrad korrigierend betätigte, sobald das Fahrzeug im Begriff war, von der Spur abzukommen.[53]

Für ein juristisches Anwendungsbeispiel kann der obige Datensatz mit zusätzlichen Daten bestehend aus Absätzen von Entscheidungsgründen, die sich nicht mit Fahrlässigkeitsstrafe beschäftigen zu einem erweiterten Datensatz angereichert werden. Ein Reinforcement Learning-Algorithmus könnte nun die Aufgabe zu bewältigen haben, die Daten aus dem erweiterten Datensatz als zum Datensatz gehörig oder nicht gehörig zu klassifizieren. Jede richtige Klassifikation würde dem Algorithmus als „Reward" zurückgemeldet werden; jede falsche nicht.

3. Künstliche neuronale Netzwerke

Es existieren unzählige Algorithmen, mit denen Maschinelles Lernen betrieben werden kann, und der Bestand wächst stetig. [54] Viele Algorithmen lassen sich bestimmten Lerntypen zuordnen, weil sie typischerweise nur für solche in Frage kommen. So kommen für eine Klassifikationsaufgabe im Supervised Learning etwa Entscheidungsbäume (Decision Trees), künstliche neuronale Netzwerke (Artificial

51 *Russel/Norvig*, AI (Fn. 23), S. 830.

52 *A. Kendall et al.*, Learning to Drive in a Day, 2018, arXiv:1807.00412.

53 *Kendall et al.*, Learning to Drive (Fn. 52), S. 3.

54 Ein erstes Gefühl für die sehr große Menge an existierenden Modellen und Algorithmen liefern die Anwendungsbeispiele von scikit-learn (Fn. 35), S. 637 ff.

Neural Networks), k-Nächste-Nachbarn (k-Nearest-Neighbours), Support Vektor Maschinen (Support Vector Machines) oder Bayessche Netze (Bayesian Networks) als Algorithmen in Betracht.[55] Für das Unsupervised Learning existieren z.B. diverse Clustering-Algorithmen, aber es kommen auch künstliche neuronale Netzwerke zum Einsatz.[56] Schließlich spielen beim Reinforcement Learning insbesondere Markow-Entscheidungsprobleme (Markov Decision Process) eine wichtige Rolle.[57] All diesen Fällen ist gemein, dass im trainierten Algorithmus das Entscheidungs- und Aktionswissen des Agenten enthalten ist.

An dieser Stelle soll exemplarisch das Maschinelle Lernen unter Einsatz sog. künstlicher neuronaler Netzwerke ein wenig detaillierter betrachtet werden. Dies geschieht einerseits, weil im Zusammenhang mit ihnen besonders häufig von „Autonomie", „Künstlicher Intelligenz" und „Black Box" die Rede ist, [58] andererseits, weil die jüngsten Durchbrüche in der KI-Forschung auf ihrem Einsatz beruhen, [59] und schließlich, weil sie im Rahmen aller Lerntypen zum Einsatz kommen können.[60]

Wie die Bezeichnung „künstliches neuronales Netzwerk" bereits nahelegt, wurden diese Algorithmen erfunden, um die Funktionsweise biologischer Neuronen mathematisch zu modellieren.[61] Das künstliche neuronale Netzwerk besteht, im übertragenen Sinne gesprochen, aus einer Mehrzahl von Neuronen, welche die Knoten des Netzwerkes bilden und durch Signal-übertragende Kanten verbunden werden. Dabei kann ein Neuron mehrere Input-Signale aufnehmen, die jeweils ein bestimmtes Gewicht (Weight) aufweisen. Das Gewicht kann als Übertragungsstärke interpretiert werden. Das Neuron selbst wird über eine Aktivierungsfunktion (Activation

55 In Überblick *S. Kotsiantis*, Supervised Machine Learning: A Review of Classification Techniques, Informatica 2007, S. 249 (251 ff.).

56 scikit-learn (Fn. 35), S. 322 ff., 396 ff.

57 *Russel/Norvig*, AI (Fn. 23), 645 ff., 830 ff.

58 Zur Bezeichnung als "Black Box" bereits *J. Benitez/J. Castro/I. Requena*, Are artificial neural networks black boxes?, IEEE Transactions on Neural Networks 1997, 1156 ff.

59 Man denke etwa an AlphaGo, das 2016 erstmalig einen der weltweit besten Go-Spieler besiegen konnte, vgl. *D. Silver et al.*, Mastering the game of Go with deep neural networks and tree search, Nature 2016, S. 484 ff.; oder an IBMs Debater Projekt, das mehr als nur auf Augenhöhe mit Menschen debattieren kann, vgl. https://www.research.ibm.com/artificial-intelligence/project-debater/research.html, zuletzt abgerufen am 15.2.2019.

60 Zu den jeweiligen Lerntypen unter Einsatz künstlicher neuronaler Netzwerke, s. *J. Schmidhuber*, Deep Learning in Neural Networks: An Overview, IDSIA Technical Reports 2014; z.B. wurde AlphaGo mit künstlichen neuronalen Netzwerken in einer Kombination von Semi-Supvervised Learning und Reinforcement Learning trainiert (Fn. 59), während die Weiterentwicklung AlphaGo Zero nur auf Reinforcement Learning mit künstlichen neuronalen Netzwerken beruhte, s. *D. Silver et al.*, Mastering the game of Go without human knowledge, Nature 2017, S. 354 ff.

61 Erstmals *W. McCulloch/W. Pitts*, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 1943, S. 115 ff.; seitdem hat sich die Modellierung biologischer neuronaler Systeme wesentlich weiterentwickelt, während künstliche neuronale Netzwerke als Gegenstand mathematischer und statistischer Forschung interessant blieben, vgl. *Russel/Norvig*, AI (Fn. 23), S. 728.

Function) gesteuert, die einen variablen Bias aufweist, wodurch bestimmt wird, wie „stark" das Übertragungssignal insgesamt sein muss bis es „feuert" und damit einen Output erzeugt.[62]

Das einfachste künstliche neuronale Netzwerk ist das sog. Perzeptron bzw. Perzeptron-Netzwerk. Es hat die Eigenschaft, dass die Inputs der Netzwerk-bildenden Neuronen direkt mit dem Output verbunden sind, sodass ein Netzwerk mit nur einer „Schicht" (Single-Layer) von Neuronen entsteht, wobei auch bereits einzelnes Neuron als Perzeptron bezeichnet wird.[63] Die Inputs stellen dabei die Features des Modells dar, während ein Output das jeweilige Label darstellt.[64] Mit dem Perzeptron können Klassifikationsprobleme gelöst werden, wobei das Lernen darin besteht, dass das Netzwerk die jeweiligen Gewichte und den Bias durch Verarbeitung des Feedbacks im jeweiligen Lern-Setup richtig justiert.[65]

Sobald das algorithmische Modell mehr als nur ein Layer an Neuronen aufweist, ist von „Deep Learning" die Rede. Dabei bezeichnet „Deep", dass das Netzwerk mehr als nur über ein Input-Layer und ein Output-Layer verfügt, sondern dazwischen noch ein oder mehrere sog. Hidden-Layer liegen.[66] Auch das Deep Learning kennt viel Spielarten. Zu nennen wären zunächst einfache Multilayer Perceptrons bzw. Feed-Forward Neural Networks, bei denen Input-Informationen nur in eine Richtung durch das Netzwerk verarbeitet werden und vom Output keine Rückführung zurück ins System erfolgt.[67] Viel Beachtung haben sog. Convolutional Neural Networks erfahren, die in der praktischen Anwendung besonders gute Vorhersagen treffen.[68] Aufgrund der vielen im Lernprozess beteiligten Neuronen und ihrer jeweiligen Weights and Biases, stellen Deep Learning Algorithmen eine große Herausforderung für die Erklärbarkeit erzielter Output-Ergebnisse dar. Es kann derzeit nur sehr schwer bis überhaupt nicht rekonstruiert werden, weshalb für bestimmte Neuronen exakt die beobachteten Weights and Biases eingestellt wurden und wie sich diese auf die Datenverarbeitung im gesamten Netzwerk auswirkt.[69]

62 *Russel/Norvig*, AI (Fn. 23), S. 727 f.

63 *Russel/Norvig*, AI (Fn. 23), S. 729 f.

64 *Hoogendoorn/Funk*, Machine Learning (Fn. 36), S. 125 f.

65 *Hoogendoorn/Funk*, Machine Learning (Fn. 36), S. 127.

66 *Russel/Norvig*, AI (Fn. 23), S. 729.

67 *I. Goodfellow/Y. Bengio/A. Courville*, Deep Learning, Cambridge, Mass.: The MIT Press 2016, S. 164.

68 *Goodfellow/Bengio/Courville*, Deep Learning (Fn. 67), S. 326 ff., *Hoogendoorn/Funk*, Machine Learning (Fn. 40), S. 129 ff.

69 Vgl. *Schweighofer et al.*, algorithmische Entscheidungsverfahren (Fn. 32), S. 54 f.; zum Lernprozess *Kotsiantis*, Supervised ML (Fn. 55), S. 255 f. und zu Analysemöglichkeiten *Adler et al.*, Auditing black-box models (Fn. 32), S. 95 ff. jeweils bezogen auf Feed-Forward Neural Networks.

4. Evaluation

Dass die Datenverarbeitung durch Machine Learning Algorithmen, insbesondere künstliche neuronale Netzwerke, teilweise schwer zu erklären ist, bedeutet natürlich nicht, dass eine Evaluation nicht möglich wäre. Ganz im Gegenteil: Eine Evaluation der Algorithmen findet stets statt, weil sonst keine Aussage darüber getroffen werden könnte, wie gut die Vorhersagekraft des jeweiligen Algorithmus ist. Für Klassifikationsaufgaben setzt z.B. eine gängige Metrik die Zahl der wahr positiven (true positive, TP), falsch positiven (false positive, FP), wahr negativen (true negative, TN) und falsch negativen Vorhersagen (false negative, FN)[70] in bestimmte Verhältnisse, woraus Zahlenwerte gebildet werden, die zwischen 0 und 1 liegen und regelmäßig umso besser sind, je näher sie sich 1 annähern.

Besonders gängig sind dabei Accuracy, Precision, Recall und Fβ-Score.[71] Accuracy bezeichnet das Verhältnis aller richtigen Klassifikationen (TP + TN) zur Summe aller vorgenommen Klassifikationen (TP + FP + TN + FN). Was intuitiv als sinnvolles Gütemaß erscheint, versagt bei Datensätzen, bei denen positive und negative Instanzen nicht ausgeglichen sind. In einem Datensatz von 100 Instanzen, der nur eine einzige TP-Instanz enthält, würde ein – unbrauchbarer – Algorithmus, der pauschal alle Instanzen als negativ klassifiziert, eine sehr gute Accuracy von 0.99 erzielen. Precision bezeichnet das Verhältnis aller TP-klassifizierten Instanzen zur Summe aller als positiv klassifizierten Instanzen (TP + FP). Der obige Algorithmus würde dazu führen, dass unzulässigerweise 0 durch 0 zu teilen wäre. Das Ergebnis ist so zu interpretieren, dass der Algorithmus keinerlei Information über positiv klassifizierte Instanzen enthält, was in diesem Fall auch einleuchtet. Hohe Precision Werte bedeuten dagegen, dass der Algorithmus wenig negative Instanzen fälschlicherweise als positiv klassifiziert. Recall bezeichnet das Verhältnis der TP-klassifizierten Instanzen zur Summe aller positiven Instanzen (TP + FN). Im gebildeten Beispiel wäre der Recall 0 geteilt durch 1 und damit 0, sodass über den Algorithmus ausgesagt werden kann, dass er überhaupt nicht im Stande ist, positive Instanzen zu erkennen. Der Fβ-Score (auch Fβ-Measure) setzt Precision und Recall in ein gewichtetes Verhältnis und ermöglicht eine bessere Interpretation der Klassifikationsgüte, wenn – wie im Beispiel – positive und negative Instanzen im Datensatz ungleich gewichtet sind.

5. Verfahren

Das Verfahren, einen maschinellen Lernalgorithmus zu trainieren, ist aufwendig und meist auf die Zusammenarbeit unterschiedlicher Experten(gruppen) angewie-

70 True positive: Anzahl der richtig als positiv klassifizierten Instanzen; false positive: Anzahl der fehlerhaft als positiv klassifizierten Instanzen; true negative: Anzahl der richtig als negativ klassifizierten Instanzen; false negative: Anzahl der falsch als negativ klassifizierten Instanzen.

71 Prägnante Zusammenfassung mit juristischen Beispielen und evaluierten Anwendungsfällen, s. *Ashley*, AI and Legal Analytics (Fn. 36), S. 113 ff.

sen. Die Vorstellung, dass „irgendwie" Daten bereitgestellt werden, und der Algorithmus „eigenverantwortlich" die Arbeit erledigt, entspricht nicht der Realität. Das beginnt schon bei der Formulierung des zu lösenden Problems (z.B. die Erledigung einer bestimmten Klassifikationsaufgabe). Anschließend ist über ein aufzustellendes Modell und die gewählten Methoden zu ermitteln, welche Daten erforderlich sind. Dann müssen Daten gesammelt oder erhoben werden, woran häufig eine Vorverarbeitung (Pre-Processing) anknüpft, um die Daten in ein Format zu überführen, das algorithmisch verarbeitbar ist. Im Falle des Supervised Learning werden die Daten nun in ein Trainings- und Test-Set aufgeteilt. Erst danach erfolgt das Training von Machine Learning Algorithmen unter Verwendung des Trainingssets. Nach dem Training wird der trainierte Algorithmus anhand der Test-Daten evaluiert.[72] Dabei können in jedem Arbeitsschritt Nachjustierungen erforderlich sein, die sich auf das gesamte Verfahren auswirken. Insgesamt kann selbst bei dieser höchst kursorischen Beschreibung festgestellt werden, dass Maschinelles Lernen von vielfältigen menschlichen Entscheidungen abhängt.

C. Fahrlässigkeitsdelikt

Wenn lernende Roboter zunehmend im menschlichen Umfeld agieren, ist es wortwörtlich „vorprogrammiert", wenngleich selten intendiert, dass Menschen geschädigt oder gar getötet werden. Die bereits eingetretenen Todesfälle im Zusammenhang mit automatisierten[73] bzw. autonomen[74] Fahrzeugen stehen dafür Beispiel.[75] Damit stellt sich die Frage strafrechtlicher Verantwortung, wobei das StGB außerhalb vorsätzlichen Handelns eine Erfolgsverursachung „durch Fahrlässigkeit" (§§ 222 und 229 StGB) voraussetzt.

Strafandrohung und strafrechtliche Sanktionen sind die schärfste Möglichkeit des Staates verhaltensregulierend zu intervenieren und sind deshalb vor dem Hinter-

72 Dazu im Überblick für eine Klassifikationsaufgabe im Supervised Learning, s. *Kotsiantis*, Supervised ML (Fn. 55), S. 250 f.; allgemein zur Implementierung von ADM-Systemen, s. *Schweighofer et al.*, algorithmische Entscheidungsverfahren (Fn. 32), S. 45 ff.; *A. Zweig*, Wo Maschinen irren können, Gütersloh 2018, S. 17 ff., abrufbar über: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/wo-maschinen-irren-koennen/, abgerufen am 15.2.2019.

73 Der erste Todesfall ereignete sich 2016 bei einem Fahrzeug, das im hochautomatisierten Fahrmodus betrieben wurde und mit einem Lastwagen kollidierte, vgl. https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk, abgerufen am 15.2.2019.

74 Der tödliche Unfall mit einem Radfahrer im Zusammenhang mit einem autonomen (Test)Fahrzeug hat 2018 als ersten Fall dieser Art viel Aufsehen erregt, vgl. https://www.nytimes.com/2018/05/24/technology/uber-autonomous-car-ntsb-investigation.html, abgerufen am 15.2.2019.

75 Allgemein zur Fahrlässigkeitshaftung im Zusammenhang mit autonomen Fahrzeugen, *S. Beck*, Selbstfahrende Kraftfahrzeuge – aktuelle Probleme der (strafrechtliche) Fahrlässigkeitshaftung, in: B. Oppermann/J. Stender-Vorwachs (Hrsg.), Autonomes Fahren – Rechtsfolgen, Rechtsprobleme, technische Grundlagen, München 2017, S. 33 ff.; auch aus US-rechtlicher Perspektive *S. Gless/E. Silverman/T. Weigend*, If Robots Cause Harm, Who Is to Blame: Self-Driving Cars and Criminal Liability, New Criminal Law Review, S. 412 ff.; und speziell Sorgfaltspflichten s. *B. Valerius*, Sorgfaltspflichten beim autonomen Fahren, in: E. Hilgendorf (Hrsg.), Autonome Systeme und neue Mobilität, S. 9 ff.

grund von Rechtsstaatsprinzip und Verhältnismäßigkeitsgrundsatz nur als ultima ratio einzusetzen.[76] Gerade im Umgang mit innovativen und gleichzeitig weitreichenden Technologien, wie lernfähigen Robotern und anderen maschinell lernenden Agenten, muss das Strafrecht mit Augenmaß operieren. Die Ausdehnung oder Schaffung am Rechtsgüterschutz ausgerichteter strafrechtlicher Verhaltensregeln schränkt Handlungsspielräume ein und kann, auch unbeabsichtigt, Innovationen behindern. Es geht einerseits um Wahrung der Freiräume zur Verwirklichung von Innovationspotenzial und anderseits um Eingrenzung von Innovationsrisiken.[77] Da die Ziehung neuer Grenzen nicht auf einem unbeschriebenen Blatt erfolgt, müssen zunächst bereits bestehende Grenzen analysiert werden, um sinnvoll entscheiden zu können, wo es überhaupt einer Grenzverschiebung oder gar neuer Grenzen bedarf. Eine als zu großzügig gesehene Verantwortungsverschiebung auf „autonom" agierende Roboter trotz vielfältiger menschlicher Entscheidungen im maschinellen Lernprozess könnte gerade dazu führen, dass übereilt neue Verhaltensregeln eingeführt werden. Deshalb soll beispielhaft anhand der fahrlässigen Schädigung und Tötung von Menschen (§§ 222, 229 StGB) untersucht werden, wie die Fahrlässigkeitsstrafdrohung bereits heute verhaltensregulierend auf Programmierhandlungen beim Maschinellen Lernen einwirkt, um im Anschluss Gedanken zur Regulierung de lege ferenda zu formulieren.

1. Erfolgsverursachung

Wenn mittels Roboter Menschen verletzt oder getötet werden, stellt die Feststellung der Erfolgsverursachung keine besondere Herausforderung dar, sofern Kausalität im Sinne der Äquivalenztheorie als „jede Handlung, die nicht hinweggedacht werden kann, ohne dass der tatbestandliche Erfolg entfiele",[78] verstanden wird. Der Programmierer, welcher den Roboter durch Auswahl und Design des Lernverfahrens konzipierte und durch Bereitstellung des Computercodes und der Trainingsdaten programmierte, hat durch sein Verhalten eine Ursache gesetzt, die für den Erfolgseintritt nicht fehlen durfte. Dabei ist der Rekurs auf „den Programmierer" natürlich deutlich unterkomplex, weil ganz in der Regel kein einzelner Programmierer allein das gesamte Maschinelle Lernverfahren konzipiert, durchführt und evaluiert, sondern große Teams unterschiedlicher Experten am Verfahren beteiligt sind. Allerdings soll weder die allgemeine Frage der Verantwortungsabgren-

76 Vgl. BVerfGE 120, 224 (238 f.); mit kritischer Reflexion auch *H. Landau*, Die deutsche Strafrechtsdogmatik zwischen Anpassung und Selbstbehauptung – Grenzkontrolle der Kriminalpolitik durch die Dogmatik?, ZStW 2009, S. 965 (971 f.).

77 Zur Bedeutung des Strafrechts für die „Risikogesellschaft", s. bereits *E. Hilgendorf*, Strafrechtliche Produzentenhaftung in der „Risikogesellschaft", Berlin 1992, S. 43 ff.

78 Bereits BGHSt 1, 332 (333), allgemein zur Kausalität im Sinne einer condicio-sine-qua-non und mit Verweis auf andere Definitionsansätze, statt vieler: *T. Fischer*, Strafgesetzbuch mit Nebengesetzen, 66. Aufl., München 2019, Vor § 13 Rn. 20 ff. mwN.

zung im Rahmen arbeitsteiliger Prozesse,[79] noch die mögliche Konstruktion einer fahrlässigen Mittäterschaft[80] oder die Nebentäterschaft bei Fahrlässigkeitsdelikten[81] Gegenstand vorliegender Betrachtung sein. Die Fiktion eines Einzelprogrammierers, der den gesamten Prozess verantwortet, dient dem konzentrierteren Blick auf typische Fragen des Fahrlässigkeitsvorwurfs.

2. Sorgfaltspflichtverletzung

Die „Fahrlässigkeit" wird in objektiver Hinsicht allgemein als objektiv vorhersehbare bzw. erkennbare Sorgfaltspflichtverletzung operationalisiert.[82] Die Sorgfaltspflichtverletzung setzt zunächst die Ermittlung und Auswahl der für den betreffenden Gegenstandsbereich anwendbaren Verhaltenspflichten voraus, die das sorgfältige Verhalten beschreiben und damit zum Beurteilungsmaßstab des Verhaltens werden. Anschließend ist zu ermitteln, ob das Verhalten der betreffenden Person negativ von diesem Maßstab abgewichen ist und damit die Verhaltenspflicht verletzt hat. Ist letzteres nicht der Fall und dennoch ein Erfolg eingetreten, hat sich in der Regel lediglich ein erlaubtes Risiko bzw. das allgemeine Lebensrisiko im Erfolg verwirklicht, was straflos bleibt.[83]

Für den Anwendungsbereich der Fahrlässigkeitsdelikte ist damit der Umfang der Verhaltensregeln entscheidend, die den Sorgfaltsmaßstab bilden. Solche Sorgfaltsregeln können sich aus Gesetzen und untergesetzlichen Rechtssätzen ergeben, wie z.B. der StVO für die Teilnahme am Straßenverkehr, aber ebenso aus nicht-staatlichen Regeln (bzw. einschränkend auf deren Basis), wie Industrienormen, etablierten Berufsgepflogenheiten oder Spielregeln im Sport. Diese können auch generalklauselartig ausfallen, wie die „allgemein anerkannten Regeln der Technik" oder „Regeln der ärztlichen Kunst" zeigen.[84] Auf derartige Beschreibungen, wie dem „Stand der Wissenschaft und Technik" (§ 1 Abs. 2 Nr. 5 ProdHG) oder dem

79 Ausführlich *M. Hannes*, Der Vertrauensgrundsatz bei arbeitsteiligem Verhalten, Aachen 2002, S. 53 ff.; im Zusammenhang mit Arzneimittelschäden, *M. Mayer*, Strafrechtliche Produktverantwortung bei Arzneimittelschäden, Berlin 2008, S. 418 ff.; allgemein zum Vertrauensgrundsatz statt vieler, *C. Roxin*, Strafrecht Allgemeiner Teil, Band 1, 4. Aufl., München 2006, § 24 Rn. 21 ff. jeweils mwN.

80 Zum Meinungsstand *Fischer*, StGB (Fn. 78), § 25 Rn. 49 ff.; *C. Roxin*, Strafrecht Allgemeiner Teil, Band 2, München 2003, § 25 Rn. 239 ff.; ausführlich *J. Böhringer*, Fahrlässige Mittäterschaft, Baden-Baden 2017, S. 80 ff.

81 BGH NJW 2010, 1087 (1092) m. Anm. *K. Kühl*; *Fischer*, StGB (Fn. 78), § 25 Rn. 53.

82 Zu den unterschiedlichen Auffassungen zur Konkretisierung des Begriffs der Fahrlässigkeit, vgl. *K. Kühl*, Strafrecht Allgemeiner Teil, 8. Aufl., München 2017, § 17 Rn. 14 ff. mwN; zum Zusammenhang zwischen zivilrechtlicher und strafrechtlicher Fahrlässigkeit, *L. Blechschmitt*, Der Fahrlässigkeitsmaßstab im Straf- und Zivilrecht am Beispiel des Einsatzes von Medizintechnik im Rahmen ärztlicher Behandlung, in: E. Hilgendorf/S. Hötitzsch (Hrsg.), Das Recht vor den Herausforderungen der modernen Technik, Baden-Baden 2015, S. 115 ff.

83 Vgl. *Kühl*, StGB AT (Fn. 82), § 17 Rn. 16 f.; zu besonderen Erfordernissen spezieller Gefahrenlagen s. BGHSt 37, 184, (189).

84 *Kühl*, StGB AT (Fn. 82), § 17 Rn. 23 f.; *Roxin*, StGB AT I (Rn. 79), § 24 Rn. 14 f. auch zum fehlenden Automatismus zwischen Regelverstoß und Sorgfaltspflichtverletzung.

„Stand der medizinischen Wissenschaft und Technik" (z.B. § 5 Abs. 1 TFG), rekurriert auch der Gesetzgeber. Sind spezielle Verhaltensregeln nicht vorhanden, muss in Anlehnung an § 276 Abs. 2 BGB auf die „im Verkehr erforderliche Sorgfalt" und das vorgestellte Verhalten eines „besonnenen und gewissenhaften Menschen aus dem Verkehrskreis des Täters" zurückgegriffen werden.[85] Diese Bezugnahme der Rechtsprechung auf Verkehrskreise verdeutlicht, was die unterschiedlichen Sorgfaltsregeln für unterschiedliche Situationen vorzeichnen: Die Sorgfaltsanforderungen erfahren eine personenbezogene Typisierung, die von der sozialen Rolle des Handelnden und dem jeweiligen Handlungszusammenhang abhängen.[86] Für das Maschinelle Lernen stellt sich deshalb die Frage, wie sich ein sorgfältiger Programmierer oder Software-Developer, bzw. noch granularer typisiert, wie sich ein sorgfältiger „Data Scientist", „Machine Learning Scientist", „Machine Learning Modeler" oder „Machine Learning Engineer" verhalten hätte.[87]

Gerade aufgrund der relativ kurzen und dynamischen Entwicklungsgeschichte des Maschinellen Lernens, insbesondere unter Einsatz künstlicher neuronaler Netzwerke, haben sich bisher weder gesetzliche, noch durch Industrieverbände etablierte Verhaltensstandards ausgebildet. Ein etwa für den medizinischen Bereich mit dem Paul-Ehrlich-Institut[88] vergleichbarem Bundesinstitut für Maschinelles Lernen, das Standards diskutiert und etabliert, existiert nicht.[89] Große Softwareunternehmen haben allerdings bereits angefangen, Best Practices für Machine Learning zu formulieren, welche erste Anhaltspunkte für sorgfältiges Verhalten liefern.[90]

So komplex die Implementierung eines Maschinellen Lernprozesses in der Robotik ist, so vielfältig gestalten sich auch denkbare Sorgfaltspflichtverletzungen. Um in Anknüpfung an die obige Skizze des Maschinellen Lernens einige Beispiele zu bilden: Schon bei der Modellierung des Problems können Fehler auftreten. Wer beispielsweise zur Modellierung eines autonomen Fahrzeugs nicht beachtet, dass Stop-Schilder (oder die Regeln der StVO überhaupt) Teil des Gegenstandsbereichs

85 Statt vieler *R. Rengier*, Strafrecht Allgemeiner Teil, 10. Aufl., München 2018, § 52 Rn. 15, 18. Dass solche Formulierungen zwar als schöne Floskel, aber wenig als erkennbare und einhaltbare Verhaltensregeln taugen, liegt auf der Hand. Letztlich geht es um die Einhaltung einer Art „gesundem Menschenverstand", den der Richter im Verfahren häufig nur durch den eigenen Richterverstand ersetzen kann.

86 Vgl. BGH NStZ 2005, 446 (447); *Kühl*, StGB AT (Fn. 82), § 17 Rn. 25 f.

87 Zur Berücksichtigung von Sonderwissen und Sonderfähigkeiten: *Rengier*, StGB (Fn. 85), § 52 Rn. 19 ff. mwN.

88 Zu den Aufgaben des Paul-Ehrlich-Instituts vgl. https://www.pei.de/DE/institut/aufgaben/aufgaben-node.html, zuletzt abgerufen am 15.2.2019.

89 Vereinzelt finden sich akademische Publikationen, die sich dieser Aufgabe annehmen bzw. annähern; für das Deep Learning vgl. bereits *P. Simard/D. Steinkraus/J. Platt*, Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis, ICDAR Proceedings 2003.

90 Vgl. z.B. von Google https://developers.google.com/machine-learning/guides/rules-of-ml/ oder Amazon https://aws.amazon.com/de/partners/navigate/machine-learning/, zuletzt abgerufen am 15.2.2019.

sind, vom Fahrzeug erkannt werden müssen und ein entsprechendes Haltemanöver erforderlich ist, handelt sorgfaltspflichtwidrig (Modellierungsfehler). Ebenso wäre bei der Auswahl der Trainingsdaten die Sorgfaltspflicht verletzt, wenn ein künstliches neuronales Netzwerk, das PKW von LKW unterscheiden soll, mit Daten trainiert wird, auf denen PKW stets bei klarem Himmel und LKW stets bei bewölktem Himmel abgebildet werden, sodass der Algorithmus lediglich wolkenlose und bewölkte Wetterlagen zu unterscheiden lernt (Trainingsfehler bei der Datenauswahl).[91] Im Rahmen des Supervised Learning können Sorgfaltspflichtverstöße z.B. darin bestehen, dass Menschen bei der Erstellung der Trainingsdaten diese fehlerhaft labeln (Trainingsfehler bei der Datenerzeugung). Im Rahmen der Evaluation sind ferner, wie im Falle der statistischen Signifikanz,[92] festgelegte Grenzwerte bzgl. Accuracy, Precision, Recall und Fβ-score vorstellbar, deren Unterschreiten eine Sorgfaltspflichtverletzung indizieren.[93] Dies würde dazu führen, dass bestimmte Lerntypen oder Lernalgorithmen nach der Evaluation als ungeeignet zu qualifizieren sind. Liefert kein einziger Lernalgorithmus hinreichend präzise Ergebnisse, muss von maschinellen Lernverfahren – zumindest auf der bestehenden Datengrundlage – gänzlich abgesehen werden. Zudem wird für die jeweiligen Arbeitsschritte auch die Übernahmefahrlässigkeit Relevanz gewinnen, wenn einzelne Programmierer Aufgaben im Machine Learning Prozess wahrnehmen, für die es ihnen nachweislich an der entsprechenden Expertise mangelt.[94] Kurzum: Bei der Konstruktion lernender Roboter liefert schon allein der softwareseitige maschinelle Lernprozess mannigfaltige Anknüpfungspunkte für menschliche Sorgfaltspflichtverletzungen.[95]

91 Im Einzelfall wird die Sachlage wahrscheinlich weniger deutlich sein. So war der erste Todesfall im Zusammenhang mit hochautomatisiertem Fahren (s. Fn. 73) darauf zurückzuführen, dass der Algorithmus eine helle und möglicherweise spiegelnde LKW-Plane als Himmel klassifizierte und deshalb ungebremst mit dem LKW kollidierte. Nun kann darüber gestritten werden, ob ein sorgfältiger Programmierer dies hätte vorhersehen und den Algorithmus auch mit spiegelnden Fahrzeugen hätte trainieren müssen und ob bei einem erfolgten Training der Erfolg mit an Sicherheit grenzender Wahrscheinlichkeit ausgeblieben wäre.

92 Zum Signifikanzniveau und üblichen Werten vgl. *L. Fahrmeir et al.*, Statistik, 8. Aufl., Berlin 2016, S. 374 ff.

93 So hat sich der tödliche Unfall eines autonomen Fahrzeuges mit einem Radfahrer (s. Fn. 74) deshalb ereignet, weil die beteiligten Programmierer den Notbremsmechanismus ausgeschaltet haben, damit Testfahrten weniger „stockend" ablaufen können.

94 Im Zusammenhang mit ärztlichen Tätigkeiten, vgl. BGHSt 43, 306 (311); 55, 121 (133); 56, 277 (287); zur Übernahmefahrlässigkeit im Allgemeinen, s. *G. Duttge*, in: Münchener Kommentar zum Strafgesetzbuch, Band 1, 3. Aufl., München 2017, § 15 Rn. 131 ff.; *Kühl*, StGB AT (Fn. 82), § 17 Rn. 35 jeweils mwN.

95 Und während des Betriebs autonomer Roboter trifft den Produzenten oder Betreiber natürlich produktbezogene Verkehrssicherungspflichten, vgl. *G. Timpe*, Die strafrechtliche Produzentenhaftung, HRRS 2017, S. 272 (273 ff.); *P. Kröger*, in: H. Laufhütte/R. Rissing-van Saan/K. Tiedemann (Hrsg.), Strafgesetzbuch Leipziger Kommentar, Band 7 Teil 1, 12. Aufl., Berlin 2019, § 222 Rn. 35.

3. Vorhersehbarkeit

Fahrlässigkeit erfordert ferner, dass im Zeitpunkt der Sorgfaltspflichtverletzung der Eintritt des Erfolgs vorhersehbar bzw. erkennbar war.[96] In objektiver Hinsicht verlangt Vorhersehbarkeit nicht, dass der mögliche Täter den Erfolg in allen Einzelheiten voraussehen konnte, sondern es genügt, wenn die Auswirkungen der Sorgfaltspflichtverletzung in Art, Umfang und Ausmaß erkennbar waren.[97] Im Zusammenhang mit Robotern, die typischerweise mit Menschen in Kontakt kommen, wie autonome Fahrzeuge oder Serviceroboter wird die Vorhersehbarkeit regelmäßig vorliegen, wenn allein darauf abgestellt wird, dass diese Roboter ihrer Bestimmung entsprechend mit Menschen interagieren und aufgrund ihrer physikalischen Eigenschaften und durch ihr Aktionsprogramm Menschen verletzen können. Dies betrifft insbesondere Programmierfehler im Zusammenhang mit der Befolgung von Straßenverkehrsvorschriften, welche gerade der Unfallverhütung dienen.[98] Ein eher weites Verständnis der Vorhersehbarkeit scheint auch dadurch gedeckt, dass die Rechtsprechung im Zusammenhang mit besonderen Opfereigenschaften, wie z.B. relativ seltenen Fällen herabgesetzter Blutgerinnungsfähigkeiten, keinen Ausschluss der Vorhersehbarkeit annimmt.[99] Insoweit ist es auch unerheblich, dass z.B. im Rahmen des Deep Learning nicht erklärt werden kann, weshalb sich im Netzwerk einzelne Parameter so und nicht anders eingestellt haben. Die technische Blackbox darf auch juristisch eine solche bleiben, weil der Rechtsbegriff der Vorhersehbarkeit eine detailliertere Analyse nicht erfordert.

4. Zurechnung

Die wertungsmäßige Zuschreibung des Erfolgseintritts als Folge der Sorgfaltspflichtverletzung an den Täter im Sinne einer objektiven Zurechnung findet beim Fahrlässigkeitsdelikt einerseits allgemein als normative Einschränkung der im Sinne der Äquivalenztheorie sehr weit verstandenen Kausalität statt, und andererseits muss der Erfolg gerade „durch" Fahrlässigkeit (§§ 222 und 229 StGB) herbeigeführt worden sein, was als normtextlicher Anknüpfungspunkt für spezielle Zurechnungsfragen bei diesen Fahrlässigkeitsdelikten gesehen werden kann.[100] Die Recht-

96 Vgl. mit kritischer Analyse *Kühl*, StGB AT (Fn. 82), § 17 Rn. 18 f. mwN; oder *H. Frister*, Strafrecht Allgemeiner Teil, 8. Aufl., München 2018, 12. Kap. Rn. 1 ff., welcher Fahrlässigkeit allein als „Erkennbarkeit" versteht.

97 So etwa BGHSt 12, 75 (77); 37, 179 (180); 39, 322 (324); es wird die Einschätzung vertreten, dass die Rechtsprechung das Merkmal der Vorhersehbarkeit derart verwendet, um in bestimmten Einzelfällen gezielt von der Fahrlässigkeitsstrafe abzusehen, s. *Krüger*, LK (Fn. 95), § 222 Rn. 36.

98 Vgl. *Kröger*, LK (Fn. 95), § 222 Rn. 38.

99 RGSt 54, 349; *Kröger*, LK (Fn. 95), § 222 Rn. 39.

100 Allgemein zur objektiven Zurechnung, *T. Walter*, in: H. Laufhütte/R. Rissing-van Saan/K. Tiedemann (Hrsg.), Strafgesetzbuch Leipziger Kommentar, Band 1, 12. Aufl., Berlin 2007, Vor § 13 ff. Rn. 89 ff.; kritisch dazu in letzter Zeit, s. *K. Gössel*, Objektive Zurechnung und Kausalität, GA 2015, S. 18 ff.; und vgl. *W. Frisch*, Erfolgsgeschichte und Kritik der objektiven Zurechnungslehre, GA 2018, S. 553 ff.; *R. Planas*, Die "Lehre von der objektiven Zurechnung": Gedanken über ihren Ursprung und ihre Zukunft, GA 2016, S. 284 (287 ff.).

sprechung hat sich mit dem Begriff der „objektiven Zurechnung" noch nicht anfreunden können, folgt ihm aber gedanklich und spricht bisweilen vom „ursächliche[n] Zusammenhang"[101] oder vom Pflichtwidrigkeitszusammenhang,[102] wenn sie die objektive Zurechnung behandelt. Im Allgemeinen setzt die objektive Zurechnung die Schaffung eines rechtlich missbilligten Risikos voraus, das sich im tatbestandsmäßigen Erfolg verwirklicht.[103] Aufgrund dieses weitreichenden Verständnisses der objektiven Zurechnung wird teilweise auch vertreten, dass es für das Fahrlässigkeitsdelikt gar keiner objektiv vorhersehbaren Sorgfaltspflichtverletzung bedarf, weil die Sorgfaltspflichtverletzung mit der Schaffung eines rechtlich missbilligten Risikos identisch und bei mangelnder Vorhersehbarkeit des Erfolgseintritts gleichsam kein rechtlich missbilligtes Risiko geschaffen worden sei.[104] Dieser Gedanke erscheint schlüssig, schadet aber der hiesigen Beurteilung der Fahrlässigkeitsstrafbarkeit nicht, weil dadurch die bisher dargestellten Voraussetzungen nicht modifiziert werden. Vielmehr besteht nun die Chance, spezifischen Aspekten der objektiven Zurechnung beim Fahrlässigkeitsdelikt im Zusammenhang mit lernenden Robotern besondere Beachtung zu schenken.

a) Pflichtwidrigkeitszusammenhang

Rechtsprechung und weite Teile der Literatur verlangen für den Pflichtwidrigkeitszusammenhang zwischen Fahrlässigkeit und Taterfolg, dass der Erfolg bei einem vorgestellten pflichtgemäßen Verhalten mit an Sicherheit grenzender Wahrscheinlichkeit ausgeblieben wäre.[105] Ansonsten hätte der in Frage stehende Fahrlässigkeitstäter auch bei pflichtgemäßem Alternativverhalten den Erfolg nicht verhindern können, sodass nicht angenommen werden kann, dass das der Sorgfaltspflichtverletzung innewohnende Risiko sich im Erfolg verwirklicht hat.[106] Im Prozess hat dies zur Folge, dass in dubio pro reo der Pflichtwidrigkeitszusammenhang verneint wird, wenn die Möglichkeit nicht auszuschließen ist, dass der Erfolg auch bei pflichtgemäßem Verhalten eingetreten wäre.[107] Dieser Maßstab verlangt dem Gericht die Rekonstruktion eines hypothetischen Kausalverlaufs unter Annahme eines pflichtgemäßen Verhaltens ab, bei dem am Ende hinreichender Sicherheit ausgesagt

101 BGHSt 30, 228 (230); BGH NJW 1991, 501 (503).

102 BGHSt 37, 106 (116); OLG Köln NStZ-RR 2002, 304.

103 Anstelle vieler *Kühl*, StGB (Fn. 82), § 4 Rn. 43 mwN; für eine systematische der objektiven Zurechnung, s. *I. Puppe*, Das System der objektiven Zurechnung, GA 2015, S. 203 ff.

104 *Puppe*, objektive Zurechnung (Fn. 103), S. 209 ff.; *Roxin*, StGB AT I (Rn. 79), § 24 Rn. 12 f.

105 Zu dieser sog. „Vermeidbarkeitstheorie" vgl. BGHSt 11, 1 (7); 33, 61 (63 f.); stellvertretend aus der Literatur *Rengier*, StGB (Fn. 85), § 52 Rn. 26 ff.; zum Streitstand vgl. *Kühl*, StGB (Fn. 82), § 17 Rn. 51 ff. mwN; kann hingegen festgestellt werden, dass auch ein pflichtgemäßes Alternativverhalten mit Sicherheit den Erfolg herbeigeführt hätte, besteht über den Ausschluss des Pflichtwidrigkeitszusammenhangs Einigkeit, vgl. *Roxin*, StGB AT I (Rn. 79), § 11 Rn. 88 f.

106 Vgl. *Rengier*, StGB (Fn. 85), § 52 Rn. 33 mwN.

107 BGHSt 33, 61 (63); *Rengier*, StGB (Fn. 85), § 52 Rn. 2.

werden muss, dass der Erfolg dann nicht eingetreten wäre.[108] Aufgrund dieser strengen Anforderungen lässt ein beachtlicher Teil der Literatur es genügen, wenn das sorgfaltswidrige Verhalten bereits das Risiko das Erfolgseintritts erhöht hat.[109] Dieser Auffassung nach muss das Gericht die Feststellung treffen, dass das sorgfaltspflichtwidrige Verhalten im Vergleich zum sorgfältigen Verhalten das Risiko des Erfolgseintritts erhöht hat, was auch die Urteilsbildung anhand hypothetischer Verhaltensweisen erfordert.[110]

Obgleich man meinen könnte, das Maschinelle Lernen füge dem Sachverhalt eine weitere Dimension der Komplexität hinzu und erschwere damit auch die Feststellung des Pflichtwidrigkeitszusammenhangs, wird indes wohl das Gegenteil der Fall sein. Natürlich wird die Sachverhaltsermittlung im Zusammenhang mit lernenden Robotern nicht mehr ohne technischen Sachverstand möglich sein. Allerdings liefert das digitalisierte Umfeld ganz andere Möglichkeiten der Rekonstruktion vergangener Ereignisse als es heutige Zeugenaussagen oder Sachverständigengutachten erlauben. Lernende Roboter erfassen nicht nur mit ihren Sensoren das Umfeld, sondern dokumentieren regelmäßig auch, welche Daten sie erfasst haben und welche Aktion die jeweilige Umgebungssituation zur Folge hatte.[111] Wenn über diese Aktionsdaten Protokoll geführt wird, können diese zu Beweiszwecken beschlagnahmt werden,[112] was dann unter Umständen einen deutlich präziseren Blick in die Vergangenheit erlaubt. Zur Berücksichtigung des erforderlichen hypothetischen Kausalverlaufs ist in vielen Situationen vorstellbar, Computersimulationen unter Annahme eines sorgfaltspflichtgemäßen Verhaltens durchzuführen.[113] Diese würden Aufschluss darüber geben, wie der Roboter im pflichtgemäß programmierten bzw. trainierten Alternativfall agiert hätte, bzw. zeigen, mit welcher in Wahrschein-

108 Zur hypothetischen Kausalverläufen beim Fahrlässigkeitsdelikt, s. *V. Haas*, Die Bedeutung hypothetischer Kausalverläufe für die Tat und ihre strafrechtliche Würdigung, GA 2015, S. 86 (90 ff.).

109 Zur von Claus Roxin begründeten sog. Risikoerhöhungslehre, s. ausführlich *Roxin*, Grundlagenprobleme (Fn. 25), S. 168 ff., *ders.*, StGB AT I (Rn. 79), § 11 Rn. 90 f. mwN.

110 Dieser fast schon geschichtsträchtige Streit kann an dieser Stelle nicht bedeutungsgerecht beleuchtet werden. Es bleibt jedenfalls die Erkenntnis, dass beide Lösungswege Schwierigkeiten und Unsicherheiten bergen, da jeweils ein hypothetischer Kausalverlauf vorgestellt und Risikoeinschätzungen getroffen werden müssen, die auf menschlichen Verhaltensweisen beruhen, wofür nur mehr oder minder zuverlässige Modelle existieren, denen es allerdings weit überwiegend an einem Maßstab fehlt. Eine solche Messung scheint aber erforderlich zu sein, wenn es ernsthaft darum gehen soll, besonders hohe Eintrittswahrscheinlichkeiten oder Risikoerhöhungen nachzuweisen.

111 Auch unter Einsatz künstlicher neuronaler Netzwerke ist bekannt, auf Grundlage welchen Inputs welcher Output folgt. Probleme verursacht nur die Erklärung des „Warum", wenn Deep Learning und damit Hidden-Layer zum Einsatz kommen.

112 Grundlegend zur Möglichkeit der Beschlagnahme von Datenträgern und der darauf befindlichen Daten als „Gegenstände" iSd § 94 StPO, s. BVerfGE 113, 29 ff. Natürlich besteht in praktischer Hinsicht die Komplikation, dass diese Daten nicht zwingend im fraglichen Roboter selbst, sondern „in der Cloud" auf Servern gespeichert werden, die außerhalb Deutschlands oder der EU betrieben werden.

113 Beispiel zum Einsatz von Simulationen im Zusammenhang mit autonomen Fahrzeugen, s. *M. Bansal/A. Krizhevsky/A. Ogale*, ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst, 2018, arXiv:1812.03079.

lichkeit der Erfolg hätte vermieden werden können. Damit würde gerade digitale Systeme die Möglichkeit schaffen, zumindest roboterseitig den Pflichtwidrigkeitszusammenhang besser zu beleuchten, und zwar unabhängig davon, ob man dem Rechtsprechungsansatz oder der Risikoerhöhungslehre folgt.

Gerade bei der Feststellung des Pflichtwidrigkeitszusammenhangs darf der lernende Roboter nicht vorschnell pauschal als Blackbox qualifiziert werden, wenn man damit die Vorstellung verbindet, dass überhaupt keine Aussage darüber zulässig wäre, wie er bei einer vorgestellten sorgfaltspflichtgemäßen Programmierung agiert hätte. Liegt ein Modellierungsfehler vor und war schon zur Tatzeit eine geeignetere Modellierung möglich, lässt sich regelmäßig auch feststellen, welche Aktionen der Roboter bei einem sorgfaltspflichtgemäß erstellten Modell ausgeführt hätte. Trainingsfehler können dagegen eine größere Herausforderung darstellen, soweit sie nicht offensichtlich waren,[114] sondern z.B. Trainingsdaten nur ein sehr schwer zu erkennendes, problemunabhängiges Muster aufweisen oder nur ein Teil der Trainingsdaten falsch gelabelt wurde. Fehler im Zusammenhang mit der Evaluation können dagegen wohl regelmäßig leicht festgestellt werden, sofern man sich darüber einig wird, wie ein sorgfältiges Verhalten zu quantifizieren ist. Tatsächlich konnte auch bei der viel beachteten tödlichen Kollision eines autonomen Fahrzeugs mit einem Radfahrer im März 2018[115] ermittelt werden, dass das Fahrzeug den Radfahrer sensorisch als solchen erfasste, das Bremsmanöver allerdings nicht einleitete, weil Programmierer zuvor zu anderen Testzwecken den automatischen Notbremsmechanismus ausgeschaltet hatten.[116]

b) Schutzzweckzusammenhang

Die objektive Zurechnung erfordert ferner, dass zwischen Taterfolg und Sorgfaltspflichtverletzung ein Schutzzweckzusammenhang besteht. Dafür muss durch Auslegung ermittelt werden, ob die verletzte Sorgfaltsnorm gerade dem Schutz des in Frage stehenden Rechtsguts vor dem erzeugten Risiko dient.[117] Gerade in diesem Zusammenhang wird die Abhängigkeit des Fahrlässigkeitsdelikts von der Existenz möglichst expliziter Sorgfaltsregeln deutlich. Dabei würde die Auslegung deutlich vereinfacht, wenn die zu etablierenden Verhaltensregeln bzw. bestimmte Regelbereiche ausdrücklich ihre Zwecksetzung angeben würden, wie es z.B. in den Erwägungsgründen der EU-Gesetzgebung häufig der Fall ist. Ungeachtet dessen darf bereits an dieser Stelle die vorsichtige Einschätzung geäußert werden, dass Verhal-

114 Siehe obige Beispiele unter Teil C., 2.

115 S. Fn. 74.

116 Vgl. S. 3 der vorläufigen Berichts über diesen Unfall, abrufbar: https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf, zuletzt abgerufen am 15.2.2019.

117 Statt vieler vgl. *Rengier*, StGB (Fn. 85), § 52 Rn. 37 ff.; ausführlich *W. Degener*, „Die Lehre vom Schutzzweckzusammenhang der Norm" und die strafgesetzlichen Erfolgsdelikte, Berlin 2001, S. 147 ff.

tensregeln, welche die inhaltliche Gestaltung von Modellbildung, Training und Evaluation betreffen, bestimmungsgemäß sich unmittelbar auf die Aktionen des Roboters auswirken, sodass sie regelmäßig auch den Schutz des Aktionsumfelds dienen. Sofern Menschen in einer Weise Teil dieses Aktionsumfeld sind, die Verletzungen nicht fernliegend erscheinen lässt, erstreckt sich dieser Schutz auch ihr Leben und ihre Gesundheit. Dokumentations- und Transparenzpflichten würden dagegen häufig wohl nur der Sicherung von Beweismöglichkeiten oder anderer Interessen dienen.

c) Dazwischentreten eines Roboters

Wenngleich die allgemeine Frage der Abgrenzung von Verantwortungsbereichen, insbesondere auch bei arbeitsteiligem Zusammenwirken,[118] an dieser Stelle nicht dargestellt werden kann, bedarf es doch noch eines Blicks auf die Frage, ob durch ein „Dazwischentreten eines Roboters" die Zurechnung an den Programmierer auszuschließen ist. Solange wir entweder technisch (noch?) nicht dazu in der Lage sind, im menschlichen Sinne autonome Roboter zu bauen oder uns normativ nicht dazu entschließen, Roboteraktionen mit menschlichem Verhalten gleichzusetzen, ist die Frage eindeutig mit „Nein" zu beantworten. Auch wenn bei uns Menschen Zweifel bestehen, ob und zu welchem Grade unser Verhalten determiniert ist,[119] sind derzeitig und mittelfristig existierende Roboter durch ihren Computercode determiniert.[120] Dass dieser Code teilweise im Zuge des Trainings Änderungen unterliegt, die sich auf die Aktionen des Roboters auswirken, ändert daran nichts. In Teil B wurde dargelegt, wie bei aller Verschiedenheit geläufiger maschineller Lernansätze im Detail letztlich doch immer menschliches Verhalten die Aktionen des Roboters festlegt und Rahmenbedingungen setzt.[121] Auch unter Einsatz noch so komplexer künstlicher neuronaler Netzwerke, wäre der Zurechnungsausschluss nicht durch ein „eigenverantwortliches" Dazwischentreten des Roboters, sondern höchstens über einen fehlenden Pflichtwidrigkeitszusammenhang zu begründen. Diese Einschätzung betrifft auch den Fall, dass mehrere Roboter unterschiedlicher Hersteller, wie z.B. mehrere autonome Fahrzeuge, involviert sind. Das „grob fahrlässige Manöver eines dritten Roboters", der einen Menschen schädigt, indem ein anderes autonomes Fahrzeug mit dem autonomen Fahrzeug des geschädigten Insassen kollidiert, führt nicht wegen der Roboteraktion an sich zum Zurechnungsausschluss, sondern es wäre zu ermitteln, ob dem Programmierer hinter dem anderen Roboter ein grob fahrlässiges Verhalten vorgeworfen werden kann.

118 Nachweise s. Fn. 79.

119 Vgl. *Jäger*, Willensfreiheit (Fn. 8), S. 7 ff. mwN.

120 S. auch *Schuhr*, Pseudozurechnung (Fn. 21), S. 17.

121 Im Grundsatz ist dies auch ohne nähere Betrachtung der unterschiedlichen Ansätze offenkundig, da alle KI-Systeme Artefakte sind. Aber nur bei einem detaillierteren Blick ergibt sich, an welchen Stellen die Einhaltung von Sorgfaltspflichten gefordert werden kann.

5. Persönliche Vorwerfbarkeit

Schließlich setzt Fahrlässigkeit auch einen subjektiven Sorgfaltspflichtverstoß und individuelle Vorhersehbarkeit voraus. Damit ist die Frage zu beantworten, ob der mögliche Fahrlässigkeitstäter nach seinen persönlichen Fähigkeiten in der Lage war, sich sorgfaltspflichtgemäß zu verhalten und selbst den eingetretenen Erfolg hätte vorhersehen können.[122] An dieser Stelle kann der Komplexität des Machine Learning Prozesses individualisiert Rechnung gezollt werden. Auch im Zusammenhang mit dem individuellen Vorwurf wird ein Übernahmeverschulden verstärkt Bedeutung gewinnen, wenn Programmierer tätig werden, denen erforderliche Kenntnisse fehlen, wobei allerdings auch dieser Aspekt durch die Herausbildung bestimmter Ausbildungs- und Kompetenzstandards bedingt ist.[123]

D. Regulierungsgedanken

Obgleich lernende Roboter vielen Menschen als autonom agierende Maschinen erscheinen, sind ihre Aktionen das Ergebnis vielfältiger menschlicher Entscheidungen. Im Rahmen des maschinellen Lernens entscheidet ein Mensch, welches Ziel ein Roboter verfolgen soll,[124] wie die Lernumgebung (abstrakt) strukturiert ist,[125] wodurch der Roboter lernen kann[126] und mit welcher Sicherheit Entscheidungen für bestimmte Aktionen ausfallen sollen.[127] Das bestehende Fahrlässigkeitsstrafrecht knüpft an diese Verhaltensweisen an und motiviert Programmierer bereits heute zu einem sorgfältigen Verhalten bei der Implementierung von Machine Learning Verfahren. Diese Verhaltensregulierung funktioniert umso besser, je deutlicher sich Verhaltensregeln etablieren, welche die Sorgfaltspflicht konkretisieren.

Die wachsende Bedeutung, welche autonome Roboter und andere Agenten im täglichen Leben einnehmen werden, liefert Anlass zu Reflexionen über Regulierungsfragen. Dies gilt insbesondere deshalb, weil in Zukunft Aktionen ganzer Robotersysteme im Wesentlichen über eine einheitliche Programmbasis gesteuert werden können, sodass das Risikopotenzial von Programmierfehlern immens ist. Dennoch sollte nicht übersehen werden, dass das maschinelle Lernen und algorithmische Verfahren überhaupt lediglich Programmierwerkzeuge sind, mit denen vielfältige Ziele verfolgt werden können. Aus diesem Grund ist eine Regulierung von algorithmischen Entscheidungsverfahren als solche ebenso wenig zweckmäßig, wie das

122 Vgl. BGHSt 40, 341 (348); auch zur Streitfrage der Verortung der subjektiven bzw. individualisierten Anforderungen der Fahrlässigkeit, s. *Kühl*, StGB (Fn. 82), § 17 Rn. 89 ff.; *Roxin*, StGB AT I (Rn. 79), § 24 Rn. 114 ff. jeweils mwN.

123 Vgl. *Roxin*, StGB AT I (Rn. 79), § 24 Rn. 117 ff., s. ferner Fn. 94.

124 Das Ziel wird vom Problem vorgegeben, das es zu lösen gilt.

125 Dies ist Gegenstand der Modellierung.

126 Die Entscheidung für einen bestimmten Lerntyp und entsprechende Algorithmen hängt einerseits vom Problem und andererseits von den zu Verfügung stehenden Daten ab.

127 Wenn davon ausgegangen wird, dass der Roboter grundsätzlich nicht agiert, wird im Rahmen der Evaluation mitbestimmt, mit welcher Fehlertoleranz Aktionen ausgeführt werden.

allgemeine Aufstellen von Regeln zum Umgang mit Schraubenziehern, nur weil die Möglichkeit besteht, sie auch als Tötungswerkzeug einzusetzen. Es liegt auf der Hand, dass ein Hersteller harmloser Office-Software und ein Militärroboterproduzent in Bezug auf Algorithmen nicht denselben Regeln zu unterwerfen sind, nur weil beide sich maschineller Lernverfahren bedienen. Rufe nach pauschaler Algorithmenregulierung legen nur Zeugnis über ein noch ungenutztes Lernpotenzial im Algorithmenverständnis des Rufenden ab.

Für Fragen der Fahrlässigkeitsstrafe, insbesondere hinsichtlich der Aufstellung von Sorgfaltsregeln, gilt: Je größer das Risiko für die von den jeweiligen Fahrlässigkeitsdelikten geschützten Rechtsgüter ist, desto eher bedarf es expliziter Regeln für maschinelle Lernverfahren. Insoweit geht es um überhaupt nichts Neues, sondern schlicht um die Sicherung eines für erforderlich gehaltenen Niveaus des Rechtsgüterschutzes. Dabei kann auf bekannte Regelungspraktiken zurückgegriffen werden. Neben den bestehenden fahrlässigen Erfolgsdelikten ist für besonders riskante Verhaltensweisen in gefahrgeneigten Situationen auch die Neuformulierung abstrakter Gefährdungsdelikte eine Option, die den Verstoß gegen besonders wichtige Sorgfaltspflichten als solchen unter Strafe stellen.[128] Dies ist z.B. dort denkbar, wo bestimmungsgemäß mit Menschen interagierende lernende Roboter, welche aufgrund ihrer physischen Beschaffenheit Menschen erheblich schädigen können, ohne Evaluation und Prüfung ihres Aktionsprogramms vertrieben werden. Ferner erscheint es sinnvoll, bereits bestehende Zulassungs- und Akkreditierungsverfahren danach zu evaluieren, ob maschinelle Lernverfahren für den jeweiligen Gegenstandsbereich in Betracht kommen und welchen Einfluss sie haben, um entsprechende Verhaltensregeln für die Implementierung des maschinellen Lernprozesses einzuführen.[129] In diesen Zusammenhängen sollte die Formulierung von ordnungsrechtlich flankierten Dokumentations- und Transparenzpflichten besonders bedacht werden.[130] Einerseits bezweckt dies ganz allgemein die Sicherung von Beweismöglichkeiten,

128 Dazu müssen natürlich die jeweiligen Sorgfaltspflichten hinreichend bestimmt formuliert werden, was dem jeweiligen Adressaten auch die Regeleinhaltung erleichtert als die bloße Forderung, sich wie ein „besonnen und gewissenhafter" Programmierer zu verhalten. Allgemein zum Zusammenhang zwischen Rechtsgüterschutz und abstraktem Gefährdungsdelikt, s. *Roxin*, StGB AT I (Rn. 79), § 2 Rn. 68 ff.

129 Für KfZ wären z.B. neue Vorschriften im Rahmen der StVZO unter Anpassung der Bußgeldvorschrift des § 69a StVZO denkbar. Für andere Produkte bietet sich eine Anknüpfung an Art. 30 VO (EG) 765/2008 zur CE-Kennzeichnungspflicht iVm den jeweiligen produktbezogenen Richtlinien und Umsetzungsgesetzen an, insbesondere was eine Konformitätsbewertung durch die Akkreditierungsstelle betrifft. In diesem Zusammenhang bestehen bereits vielfältige Vorschriften zur Sicherung der Produktsicherheit, welche straf- und ordnungsrechtlich flankiert sind, vgl. z.B. §§ 39, 40 ProdSG für Produkte im Allgemeinen oder §§ 40-42 MPG für Medizinprodukte.

130 Transparenz meint allerdings insbesondere nicht, dass einfach nur der Computercode offengelegt werden muss. Denn dieser ist selten für Menschen nachvollziehbar, insbesondere weil bisweilen durch Einsatz sog. Uglifier Computercode bewusst für Menschen praktisch unlesbar gemacht wird, um den eigenen Programmcode zu schützen. Es geht vielmehr um die Offenlegung z.B. von Modellannahmen, eingesetzten Lernalgorithmen und der Evaluationsergebnisse.

aber andererseits in bestimmten Bereichen auch den unmittelbaren Schutz rechtlicher Interessen, wenn etwa Fragen des Datenschutzes oder der Diskriminierungsfreiheit relevant werden.[131]

E. Fazit

Lernende Roboter werden zunehmend Teil des täglichen Lebens. Ebenso wie Roboter den Umgang mit Menschen lernen, muss auch das Recht Lernprozesse im Umgang mit solchen Robotersystemen durchlaufen. Eine funktionierende „Verständigung zwischen Technik und Recht" wird maßgeblich durch ein gemeinsames Vokabular der Akteure bedingt, was wiederrum sowohl ein Verstehen des technischen Gegenstandsbereichs als auch der normativen Anforderungen des Rechts erfordert.[132] Dies setzt auf beiden Seiten eine Analyse mit deutlich höherer Granularität voraus, als dieser kursorische Blick es zu leisten vermag. Allerdings bleibt festzuhalten, dass Roboterautonomie nicht mit menschlicher Autonomie zu verwechseln, und dass Maschinelles Lernen trotz aller technologischer Begeisterung ein von menschlichem Verhalten determiniertes Werkzeug ist. Der diesen Prozess gestaltende Mensch darf im Falle der Schädigung anderer durch die Maschine nicht vorschnell durch den Verweis auf ein „eigenes autonomes Verhalten" des Roboters oder die Unvorhersehbarkeit der Aktionen solcher „Black-Boxes" aus der Verantwortung genommen werden. Das bestehende Fahrlässigkeitsstrafrecht wirkt bereits in der bestehenden Formulierung und Konkretisierung durch Rechtsprechung und Wissenschaft verhaltensregulierend. Allerdings würden explizit formulierte Sorgfaltspflichten im Umfang mit maschinellen Lernverfahren in der Robotik die Handlungssicherheit der Adressaten verbessern und damit zur Rechtssicherheit beitragen. Der Ausbau dieser Sorgfaltspflichten, aber auch die Einführung neuer Verhaltensregeln muss mit den jeweiligen Einsatzzwecken abgestimmt sein und Algorithmen dürfen als bloße Werkzeuge nicht an sich und unabhängig vom Einsatzbereich zum Anknüpfungspunkt rechtlicher Regeln gemacht werden.

131 Vgl. *M. Martini*, Algorithmen als Herausforderung für die Rechtsordnung, JZ 2017, S. 1017 (1019 ff.); *M. Martini/D. Nink*, Wenn Maschinen entscheiden … – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz, NVwZ 2017, S. 1 ff.

132 Zu den legislativen Herausforderungen der normativen Regulierung technischer Neuerungen, s. *Schuhr*, Roboter (Fn. 22), S. 230 ff.