

„Künstliche Intelligenz“, Compliance und sanktionsrechtliche Verantwortlichkeit

Von Prof. Dr. Kai Cornelius, LL.M., Heidelberg

I. Einführung

In der mehr als sechzigjährigen „Geschichte“ der Forschung zur „Künstlichen Intelligenz“¹ hat sie Höhen und Tiefen erlebt, wurde nicht nur als marginal, sondern gar als Fehlentwicklung der Informatik betrachtet.² Das hat sich geändert: Dem Einsatz von „Künstlicher Intelligenz“ wird ein enormes Potential zugeschrieben. Die von der Europäischen Kommission eingesetzte „Hochrangige Expertengruppe für Künstliche Intelligenz“ sieht in ihr ein „vielversprechendes Mittel, um das Wohlbefinden von Individuum und Gesellschaft und das Gemeinwohl zu steigern sowie zur Förderung von Fortschritt und Innovation“ beizutragen.³

So soll *vertrauenswürdige* Künstliche Intelligenz (KI) in der Lage sein, dringende gesellschaftliche Probleme wie die Überalterung, eine wachsende soziale Ungleichheit und die Umweltverschmutzung wenn nicht zu lösen, so doch zumindest zu lindern.⁴ Es lassen sich vielfältige Anwendungsbeispiele finden wie maschinelle Übersetzungsprogramme (z.B. deepl oder google translate), personalisierte Spracherkennungssoftware (z.B. Alexa, Siri oder google home) oder Systeme zum autonomen Fahren.⁵ Insbesondere der Einsatz in der medizinischen Forschung ist vielversprechend, weil KI-Technologien genutzt werden könnten, um Krankheiten zielgerichtet – auf die jeweiligen Patienten zugeschnitten – zu behandeln, aber auch, um maßgeschneiderte vorbeugende Maßnahmen anbieten zu können.⁶ Es darf aber nicht verkannt

werden, dass KI-Systeme Risiken beispielsweise „im Hinblick auf Demokratie, Rechtsstaatlichkeit, Verteilungsgerechtigkeit oder den menschlichen Geist als solchen“ bergen.⁷

Um die Vorteile von KI-Systemen zu maximieren und gleichzeitig die mit deren Einsatz einhergehenden Risiken auszuschließen bzw. zu minimieren, hat die „Hochrangige Expertengruppe“ im April 2019 die „Ethik-Leitlinien für eine vertrauenswürdige KI“ vorgestellt.⁸ Dabei geht sie davon aus, dass drei Komponenten erfüllt sein müssen: Das System muss a) rechtmäßig sein, also alle anwendbaren Gesetze und Bestimmungen einhalten, b) die Einhaltung ethischer Grundsätze und Werte garantieren, also ethisch sein und c) sowohl in technischer als auch in sozialer Hinsicht robust sein.⁹ Dies betrifft den gesamten Lebenszyklus eines KI-Systems, also von der Entwicklung (einschließlich Forschung, Entwurf, Datenbereitstellung und Erprobung) über die Einführung (einschließlich der Umsetzung) und Nutzung¹⁰ bis zur Post-nutzungsphase (wie die Umsetzung von Löschkonzepten). Obwohl sich die genannten Ethik-Leitlinien nicht explizit mit der Frage einer rechtmäßigen KI beschäftigen, spiegeln sich die Komponenten b) und c) teilweise in bestehenden Gesetzen wider bzw. sind in solche zukünftig zu integrieren.¹¹

Der nachfolgende Beitrag wendet sich auf der Grundlage dieser Ethik-Leitlinien der Frage zu, welche Risiken beim Einsatz von KI zu berücksichtigen sind, ob diese Risiken durch eine KI-Compliance eingedämmt werden können, inwieweit sanktionsrechtliche Verantwortlichkeiten betroffen

¹ Der Begriff der „Artificial Intelligence“ wurde geprägt durch Carthy/Minsky/Rochester, „A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence“, 1955, abrufbar unter

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (3.2.2020); kritisch zum Begriff und zur deutschen Übersetzung als „Künstliche Intelligenz“ Herberger, NJW 2018, 2825 (2826 f.), der lieber von „extended intelligence“ spricht.

² Kohlas, in: Savory (Hrsg.), Expertensysteme: Nutzen für Ihr Unternehmen, 2. Aufl. 1989, S. 237.

³ Hochrangige Expertengruppe für Künstliche Intelligenz, Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, abrufbar unter

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60425, S. 5 Abschnitt (9).

⁴ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 42 Abschnitt (122), mit konkreten Anwendungsbeispielen in den Abschnitten (123 ff.).

⁵ Gaede, Künstliche Intelligenz – Rechte und Strafen für Roboter?, 2019, S. 21 f.; Lenzen, Künstliche Intelligenz, 2018, S. 12 ff., 64 ff., 124 ff.

⁶ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 42 f. Abschnitte (125 f.); vgl. die dort aufgeführten Beispiele zu den Projekten REVOLVER (Repeated Evolution of Cancer,

<https://www.healtheuropa.eu/personalised-cancer-treatment/87958>, 3.2.2020), Live INCITE (personalisierte ehealth-Lösungen,

www.karolinska.se/en/live-incite, 3.2.2020), CARESSES (Roboter für die Altenpflege,

www.caresse-robot.org/en/project, 3.2.2020), MyHealth Avatar (digitale Sammlung und Zugänglichmachung von Langzeit-Gesundheitsdaten mit personalisierten Prognosen für Schlaganfall, Diabetes, Herz- Kreislauf-Erkrankungen und Bluthochdruckrisiko, www.myhealthavatar.org); vgl. auch He/Baxter/ Xu/Zhou/Zhang, Nature Medicine 25 (2019), 30 (34), mit KI-Beispielen, die bereits eine FDA-Zulassung in den USA erhalten haben.

⁷ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 2 Abschnitt (2); vgl. auch die in Abschnitt 130 ff. aufgeführten KI-Beispiele zur Identifizierung und Ortung von Personen, zum verdeckten Einsatz, Bürgerinnen- und Bürgerbewertungen und tödlichen autonomen Waffensystemen.

⁸ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 5 Abschnitt (10).

⁹ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 2 Abschnitt (1), S. 6 Abschnitt (15).

¹⁰ Vgl. Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 47 Abschnitt (147), zum Lebenszyklus von KI-Systemen.

¹¹ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 8 Abschnitt (24), S. 14 Abschnitt (49).

sind oder ob gegebenenfalls Handlungsbedarf für den Gesetzgeber besteht. Dabei geht es nicht um die Frage des Einsatzes von KI für die Lösung von Rechtsproblemen, sondern vielmehr um die durch die Nutzung von KI aufgeworfenen Rechtsprobleme.¹²

II. Künstliche Intelligenz (KI)

1. KI als Wissenschaftsdisziplin

Zunächst ist zu beachten, dass KI als wissenschaftliche Disziplin mehrere Ansätze und Techniken verfolgt wie das maschinelle Lernen („Deep Learning“ mit der Fähigkeit der KI, durch die Auswertung von Datenmassen das eigene Leistungsvermögen selbständig zu steigern),¹³ das maschinelle Denken (welches die Planung, Terminierung, Wissensrepräsentation und Schlussfolgerung sowie Suche und Optimierung betrifft) und die Robotik (welche die Steuerung, Wahrnehmung, Sensoren und Aktoren sowie weitere Techniken in cyber-physischen Systemen umfasst).¹⁴

2. Begriffsklärung

Um den Begriff der KI für die Zwecke des nachfolgenden Beitrages handhabbar zu machen, sei zunächst negativ beschrieben, dass es nicht um Systeme gehen soll, die ihre Umwelt selbständig erfassen, auf diese reagieren können und bestrebt sind, sie – ähnlich wie ein Mensch – als Grundlage der eigenen Existenz zu erhalten.¹⁵ Denn solche selbstbewussten und sich die Gründe für ihr Handeln selbst setzenden KI-Systeme existieren (derzeit) ebenso wenig wie eine KI mit Gefühlen.¹⁶ Es ist immer noch der Mensch, der das zu lösende Problem, die zu verfolgende Aufgabenstellung formuliert.¹⁷ Den Algorithmen wird als Gebilde des menschlichen Geistes ihre Rolle durch den Menschen zugewiesen.¹⁸

Das Kennzeichen von KI-Systemen ist, dass sie „in Bezug auf ein komplexes Ziel auf physischer oder digitaler Ebene agieren, indem sie ihre Umgebung durch Datenerfassung wahrnehmen, die gesammelten strukturierten oder unstrukturierten Daten interpretieren, Schlussfolgerungen daraus ziehen oder die aus diesen Daten abgeleiteten Informationen verarbeiten und über die geeignete(n) Maßnahme(n) zur Erreichung des vorgegebenen Ziels entscheiden“.¹⁹ Andererseits wird unter Betonung des Aspektes der Nachahmung

menschlichen Verhaltens schon dann von einem KI-System ausgegangen, wenn dieses menschliche Intelligenz nachahmt und Aufgaben erfüllen kann, deren Lösung bisher dem Menschen vorbehalten war.²⁰ Teilweise wird – unter Hervorhebung der Autonomie der Entscheidung des künstlichen Systems – von autonomen Softwareagenten²¹ bzw. autonom agierenden Systemen²² gesprochen. Um aus der Verwendung dieses Begriffes resultierenden etwaigen Missverständnissen vorzubeugen, ist zu betonen, dass die Autonomie dieser Systeme nur so weit geht, wie ihnen ein entsprechender Aktionsbereich durch den Menschen eingeräumt wurde.²³

Entscheidend soll nachfolgend darauf abgestellt werden, dass die intelligenten Systeme ihre Entscheidungen aufgrund eigener Erfahrungs- und Lernprozesse treffen, die ihrem Verhalten zugrunde liegenden Algorithmen selbständig ändern und damit den Weg zur Lösung (der vom Menschen gestellten) Aufgabe²⁴ selbst bestimmen können,²⁵ wobei es sich um einen graduell stetig fortschreitenden Prozess handelt.²⁶ Es geht also entscheidend darum, dass ein „künstlicher Akteur vorliegt, der seine Umwelt wahrnimmt und auf Grund der Wahrnehmungen Akte ausführen kann“.²⁷

3. Risiken

Die enge Verbindung zwischen Mensch und Maschine und die Vernetzung der Systeme untereinander führen zu einer

²⁰ Gaede (Fn. 5), S. 19.

²¹ Teubner, AcP 218 (2018), 155 (156 Fn. 1); Hengstenberg/Kirn, Rechtliche Risiken autonomer und vernetzter Systeme, 2016, S. 59; Zech, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163 (168); Gless/Weigend, ZStW 126 (2014), 561; Cornelius, MMR 2002, 353; ders., ZRP 2019, 8; dieser weite Begriff umfasst auch Roboter (hierzu Hilgendorf, in: Barton/Eschelbach/Hettinger/Kempff/Krehl/Salditt [Hrsg.], Festschrift für Thomas Fischer, 2018, S. 99 [101]; Simmler/Markwalder, ZStW 129 [2017], 20 [23], und cyberphysische Systeme, Mansdörfer, in: Barton/Eschelbach/Hettinger/Kempff/Krehl/Salditt [Hrsg.], Festschrift für Thomas Fischer, 2018, S. 155 [156]), soweit diese autonome Entscheidungen treffen können.

²² Schulz, Verantwortlichkeit bei autonom agierenden Systemen, 2015, S. 44.

²³ Herberger, NJW 2018, 2825 (2827).

²⁴ Diese wird immer vom Menschen vorgegeben, da (bisher) keine selbstbewusste KI existiert, vgl. oben die Ausführungen zu Fn. 16.

²⁵ Zech (Fn. 21), S. 171; Schulz (Fn. 22), S. 44; Yuan, RW 2018, 477 (481).

²⁶ Vgl. Silver/Hubert/Schrittwieser/Antonoglou/Lai/Guez/Lanctot/Sifre/Kumaran/Graepel/Lillicrap/Simonyan/Hassabis, Science 362 (2018), 1140, zum „Reinforcement Learning“ bei welchem vorab nur noch grundlegende Regeln und Bewertungskriterien definiert werden, bevor das System in einen Selbstlernprozess eintritt; vgl. auch Erhardt/Mona, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 61 (78 f., 92); Yuan, RW 2018, 477 (487).

²⁷ Gaede (Fn. 5), S. 21.

¹² Vgl. Herberger, NJW 2018, 2825.

¹³ Eberl, Smarte Maschinen, 2016, S. 41 ff., 91 ff.; dabei soll durch „deep“ verdeutlicht werden, dass zwischen dem Input-Layer und dem Output-Layer noch ein oder mehrere Hidden-Layer liegen; Frochte, Maschinelles Lernen, 2019, S. 174; Yuan, RW 2018, 477 (490).

¹⁴ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 47 Abschnitt (144).

¹⁵ Gaede (Fn. 5), S. 22.

¹⁶ Gaede (Fn. 5), S. 22; Herberger, NJW 2018, 2825 (2829).

¹⁷ Yuan, RW 2018, 477 (492), mit einer genaueren Beschreibung des Trainings eines Lernalgorithmus.

¹⁸ Herberger, NJW 2018, 2825 (2827).

¹⁹ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 47 Abschnitt (143).

mangelnden Transparenz.²⁸ Es werden rechtliche Verantwortungs-lücken befürchtet. Es sei nur schwer erkennbar, worauf eine Entscheidung²⁹ – und damit die Verantwortung für die Folgen derselben – zurückzuführen ist.³⁰ Um diesen Lücken zu begegnen, hat das Europäische Parlament gegenüber der Kommission die Einführung einer elektronischen Person empfohlen.³¹ Dieser Vorschlag wird von Experten in einem offenen Brief an die EU-Kommission „Artificial Intelligence and Robotics“ massiv angegriffen, da die Etablierung eines rechtlichen Status einer „Elektronischen Person“ auf einer „verzerrten Wahrnehmung“ beruhe und „unangebracht“ sei.³² Technisch würden die „aktuellen Fähigkeiten selbst der am weitesten fortgeschrittenen Roboter“ überschätzt und auch ethisch und rechtlich dürfe ein Persönlichkeitsstatus nicht vom Modell der „natürlichen Person“ abgeleitet werden, da dies sowohl der Europäischen Grundrechtecharta als auch der Europäischen Menschenrechtskonvention widersprechen würde.³³

Auch von juristischer Seite wird die Diskussion geführt, ob und inwieweit einem mit einer entsprechenden Intelligenz ausgestatteten Computersystem eine Rechtspersönlichkeit zuzusprechen ist.³⁴ Dabei wird dafür eingetreten, dass diesen „ein sorgfältig kalibrierter Rechtsstatus“ zuerkannt wird,³⁵ der autonome (KI-)Systeme durchaus als teilrechtsfähige Rechtssubjekte einordnet.³⁶ Denn nur Rechtssubjekte können Träger von Rechten und Pflichten und damit auch Adressaten von Rechtsnormen (also dem Recht unterworfen) sein.³⁷

Dies ist insoweit nachvollziehbar, als es um eine enge Verflechtung von digitalen und humanen Handlungen geht,

so dass das Ziel juristischer Haftungsnormen das Erfassen eines Gesamthandelns solcher Hybride bzw. eines umfassenden Vernetzungszusammenhangs sein sollte.³⁸ Deshalb geht es vorrangig nicht um die Frage entweder keiner oder einer vollen Rechtsfähigkeit (insbesondere durch Schaffung einer e-Person), sondern um die Entwicklung situationsadäquater Rechtskonstruktionen, die entsprechend der jeweiligen Rolle in der Mensch-Maschine-Interaktion abgestimmt sind.³⁹

Bei der rechtlichen Bewertung geht es um die Einordnung des Autonomierisikos (resultierend aus „eigenständigen Entscheidungen“ der KI-Systeme), des Verbundrisikos (resultierend aus der engen Kooperation von Mensch und KI-System) und des Vernetzungsrisikos (resultierend aus der engen Verflechtung mit anderen elektronischen Systemen).⁴⁰

a) Autonomierisiko

Dem Autonomierisiko kommt eine besondere Bedeutung zu, da dieses zu der Frage führt, ob KI-Systeme als handlungsfähige Akteure im Recht anerkannt werden.⁴¹ In diesem Zusammenhang ist es von entscheidender Bedeutung, dass keine Parallelen zur Handlungsfähigkeit menschlicher Akteure gezogen werden, sondern es zielführender ist, solche zu anderen nicht-menschlichen Akteuren (wie juristischen Personen, dem Staat oder der Kirche) zu ziehen.⁴²

Solche Gedankenkonstrukte wie das einer künstlichen Person („artificial person“)⁴³ sind im Rechtssystem nichts Ungewöhnliches.⁴⁴ Dabei werden diese Gebilde regelmäßig losgelöst von den sie konstituierenden Individuen betrachtet. So wird der Staat vom Herrscher getrennt, damit dieser nicht mit dem Tod des Menschen endet; Klöster bestehen unabhängig von Mönchen,⁴⁵ und Aktiengesellschaften unabhängig von den Vorständen. Dies verdeutlicht bereits das Vergeistigte an diesen Konstruktionen: Im Gegensatz zum Menschen, der in einem modernen Rechtsstaat mit seiner Geburt als Rechtssubjekt anerkannt wird, müssen die Gedankengebilde erst rechtlich verfasst und vom Rechtssystem anerkannt werden.⁴⁶

Die Grundlage hierfür ist, als soziales Substrat nicht eine Vielheit konkreter Menschen zu sehen, sondern den Kollektivakteur als Kette von Mitteilungen bzw. Entscheidungs-

²⁸ *Spiecker*, CR 2016, 698 (701 f.).

²⁹ Vgl. zur systemtheoretischen Sicht, die autonome Entscheidung selbst (als Handlung) in die Verantwortung zu nehmen, die Ausführungen zu Fn. 71.

³⁰ Vgl. nur *Mayinger*, Die künstliche Person, 2017, S. 213 ff.

³¹ Entschließung des Europäischen Parlaments vom 16. Februar 2017 mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103 (INL), Ziff. 59 f.), abrufbar unter <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//DE> (3.2.2020).

³² Abrufbar unter <http://www.robotics-openletter.eu/> (3.2.2020).

³³ Siehe erneut <http://www.robotics-openletter.eu/> (3.2.2020).

³⁴ *Beck*, JR 2009, 229; *Cornelius*, MMR 2002, 353 (354); *Gless*, GA 2017, 324; *Wiebe*, Die elektronische Willenserklärung, 2002, S. 80; Entschließung des Europäischen Parlaments (Fn. 31), Ziff. 59 f.

³⁵ *Teubner*, AcP 218 (2018), 155 (177).

³⁶ *Teubner*, AcP 218 (2018), 155 (178 ff.); *Schirmer*, JZ 2019, 711 (716); einen anderen Ansatz (Transformation von Maschinenarbeit in Schuldnerverhalten qua Rechtsfiktion) bevorzugend *Klingbeil*, JZ 2019, 718 (721 ff.).

³⁷ *Schuhr*, Rechtstheorie 46 (2015), 225 (251); vgl. auch *Kelsen*, Reine Rechtslehre, 1934, S. 53; *Schirmer*, JZ 2019, 711 (715).

³⁸ *Teubner*, AcP 218 (2018), 155 (162).

³⁹ Ebenso *Schirmer*, JZ 2016, 660 (663 ff.); *ders.*, JZ 2019, 711 (716 ff.); *Teubner*, AcP 218 (2018), 155 (163); *Gruber*, Bioinformationsrecht, 2015, S. 250 ff.

⁴⁰ *Teubner*, AcP 218 (2018), 155 (163 ff.; 196 ff.); *Cornelius*, ZRP 2019, 8 (9 f.).

⁴¹ Vgl. *Zech* (Fn. 21), S. 191 ff.; *Cornelius*, ZRP 2019, 8 (9).

⁴² *Teubner*, AcP 218 (2018), 155 (165); *Ortmann*, NZWiSt 2017, 241 (243 ff.).

⁴³ Vgl. nur *Hobbes*, Leviathan, 1651, Kap. XVI, wobei das Titelblatt den Staat als eigene Person symbolisiert, der aus vielen einzelnen Personen zusammengesetzt ist; *Schuhr*, Rechtstheorie 46 (2015), 225 (256 f.).

⁴⁴ *Schuhr*, Rechtstheorie 46 (2015), 225 (254).

⁴⁵ Zu diesen Beispielen *Schuhr*, Rechtstheorie 46 (2015), 225 (254).

⁴⁶ *Schuhr*, Rechtstheorie 46 (2015), 225 (255).

ketten.⁴⁷ Wenn es gelingt, entsprechende „Beiträge“ von KI-Systemen als kommunikative Ereignisse in der Form einer Einheit von Information, Mitteilung und Verstehen einzuordnen,⁴⁸ werden die Parallelen zu anderen künstlichen Entitäten deutlich.⁴⁹ Dabei sollte man nicht der Fehlvorstellung unterliegen, den Algorithmen die psychischen Kompetenzen von Menschen zuzubilligen.⁵⁰ Vielmehr kommt es für das „Verstehen“ innerhalb eines Kommunikationsprozesses zwischen Mensch und Maschine darauf an, dass die „Antwort“ des Algorithmus die „Frage“ des Menschen in der Form einer eigenen Differenz zwischen Mitteilung und Information „verstehen“ und nicht darauf, ob das Innere des Algorithmus die Äußerung eines Menschen nachvollzieht oder nicht.⁵¹

Die Teilnahme an der Kommunikation sowie Zielgerichtetheit und Mittelwahl (als intentionales Handeln) eines KI-Systems dürften zwar notwendige, aber noch nicht hinreichende Bedingungen sein, um eine rechtlich relevante Autonomie annehmen zu können. Allerdings würde es zu weit gehen, eine (irgendwie geartete – auch beschränkte) Rechtsfähigkeit erst bei einer Art digitalem Selbstbewusstsein anzunehmen.⁵² Diese Argumentation orientiert sich mehr an Ähnlichkeiten zur natürlichen Person anstatt an künstlichen Entitäten.⁵³ Dabei wird Rechtsfähigkeit mit Rechtspersönlichkeit gleichgesetzt.⁵⁴ Teubner schlägt deshalb vor, bei einer „Entscheidung unter Ungewissheit“ von einer rechtlich relevanten Autonomie auszugehen; zwischen solchen Entscheidungen ist dann der Zusammenhang zur (rechtlichen) Verantwortung (als Einstehenmüssen für Entscheidungen unter Ungewissheit) herzustellen.⁵⁵ Denn die Notwendigkeit, von einem autonomen Handeln im rechtlichen Sinne auszugehen, ist erst gegeben, wenn das KI-System (bei der Verfolgung der durch den Menschen gestellten Aufgabe) Entscheidungsvarianten hat, aus welchen es dann nach Gewichtung verschiedener Kriterien die nach seiner Programmierung und Lernleistung optimale auswählt und nicht prognostizierbare Entscheidungen trifft.⁵⁶ Die Voraussetzungen dafür sind eine Programmierung des Systems, die ihm die Möglichkeit gibt, zwischen mehreren Varianten zu entscheiden, wobei die KI selbst die Entscheidung nach einer von ihr vorgenommenen Optimierung treffen muss und auch ein Programmierer das Agieren

nicht mehr voraussagen, sondern nur noch ex post korrigieren kann.⁵⁷

Im Endeffekt ist es das Rechtssystem selbst, welches die rechtlich relevante Grenze zwischen instrumentalisiertem und autonomen Handeln festlegen muss, aber unter Zugrundelegung der wissenschaftlich erforschten Eigenständigkeit digitaler Entitäten.⁵⁸ Dabei hilft wieder ein Blick auf die formalen Organisationen: So wie es für deren rechtliche Handlungsfähigkeit nicht auf innere psychische Zustände (eine innere Denkfähigkeit) ankommt, ist dies ebenso bei KI-Systemen unerheblich; vielmehr kommt es auf die Kommunikation mit den Nutzern mitsamt den daraus erwachsenen Konsequenzen an (weshalb auch von artificial communication anstatt artificial intelligence gesprochen wird).⁵⁹

Eine weitere notwendige Voraussetzung für Autonomie ist das willentliche Handeln, wobei es auch dort nicht um einen inneren psychischen Zustand des KI-Systems geht, sondern vielmehr um die Zuschreibung zielgerichteten Handelns durch die jeweiligen Beobachter.⁶⁰ Systemtheoretisch kann so der erhöhten Komplexität, die sich einer physikalischen Beschreibung entzieht, durch ein Sozialsystem (wie dem Recht) begegnet werden, indem KI-Systemen gewisse Intentionen zugerechnet und dementsprechend Konsequenzen für die Rechtswirkungen gezogen werden.⁶¹

b) Verbundrisiko

Insbesondere das aus der engen Kooperation von Mensch und KI resultierende Verbundrisiko kann angemessen nur durch eine Berücksichtigung des Handelns im Mensch-Maschine-Verbund als Kollektivphänomen Berücksichtigung finden.⁶² Der Verbund aus Mensch und Maschine tritt mit den Dritten in eine kommunikative Beziehung, wobei auf diesen Verbund als eigenständiges Handlungssystem (als Einheit) und nicht auf den jeweils beteiligten Menschen oder Algorithmus abzustellen ist.⁶³ Dies entspricht auch der Selbstwahrnehmung einer formalen Organisation als korporativer Akteur.⁶⁴ Insofern sind die Fragen ähnlich wie bei der Verantwortlichkeit formaler Organisationen, nämlich die Nicht-Personalisierbarkeit und Unaufklärbarkeit von Kausalzusammenhängen zwi-

⁴⁷ Teubner, AcP 218 (2018), 155 (165); Ortmann, NZWiSt 2017, 241 (243 ff.).

⁴⁸ Vgl. Luhmann, Soziale Systeme: Grundriß einer allgemeinen Theorie, 1984, S. 191 ff.

⁴⁹ So Teubner, AcP 218 (2018), 155 (164).

⁵⁰ Esposito, ZfSoz 2017, 249 (250); Teubner, AcP 218 (2018), 155 (167).

⁵¹ Teubner, AcP 218 (2018), 155 (168).

⁵² Teubner, AcP 218 (2018), 155 (173).

⁵³ Vgl. Hacker, RW 2018, 243, 251; Mayinger (Fn. 30), S. 14.

⁵⁴ Schirmer, JZ 2019, 711, der diese Differenzierung prägnant herausarbeitet; vgl. aber Klingbeil, AcP 217 (2017), 848 (859).

⁵⁵ Teubner, AcP 218 (2018), 155 (174 f.).

⁵⁶ Mayinger (Fn. 30), S. 14; Teubner, AcP 218 (2018), 155 (174).

⁵⁷ Teubner, AcP 218 (2018), 155 (174).

⁵⁸ Matthias, Automaten als Träger von Rechten, 2. Aufl. 2010, S. 43 ff.; Teubner, AcP 218 (2018), 155 (170); Cornelius, ZRP 2019, 8 (9).

⁵⁹ Vgl. Teubner, AcP 218 (2018), 155 (171 f.); Esposito, ZfSoz 2017, 249 (250); Cornelius, ZRP 2019, 8 (9).

⁶⁰ Dennett, The Intentional Stance, 1987, S. 17; Matthias (Fn. 58), S. 41 ff.; Teubner, AcP 218 (2018), 155 (172); Cornelius, ZRP 2019, 8 (9).

⁶¹ Teubner, AcP 218 (2018), 155 (172); Cornelius, ZRP 2019, 8 (9).

⁶² Vgl. Teubner, AcP 218 (2018), 155 (196); Cornelius, ZRP 2019, 8 (10).

⁶³ Teubner, AcP 218 (2018), 155 (198).

⁶⁴ Ortmann, NZWiSt 2017, 241 (246).

schen Straftaten und den Entscheidungen Einzelner aufgrund diffuser innerorganisatorischer Bedingungen.⁶⁵

Deshalb kann – aus ganz ähnlichen Gründen wie bei der rechtlichen Erfassung von Verbänden – diskutiert werden, ob ein solcher Verbund als eine Zurechnungs- und Haftungseinheit angesehen werden sollte, ohne dass (wie vom individualrechtlichen Standpunkt aus gesehen) eine klare Trennung der jeweiligen Einzelhandlungen von Mensch und KI-System vorgenommen wird.⁶⁶ Eine solche kollektivrechtliche Verbundlösung ist gegenüber einer individualrechtlichen Repräsentantenlösung⁶⁷ sicherlich sehr viel weitgehender, dürfte die Realität der Mensch-Maschine-Interaktion aber dann am besten abbilden, wenn die KI-Systeme (z.B. als Chat-Bots) im Sozialleben nicht eindeutig identifizierbar teilnehmen, sondern nur in einem Interaktionszusammenhang mit anderen Akteuren wahrgenommen werden.⁶⁸

c) Vernetzungsrisiko

Besondere Herausforderungen birgt das aus der engen Verflechtung mit anderen Computern resultierende Vernetzungsrisiko. Mit der Komplexität der Vernetzungen der KI-Systeme untereinander⁶⁹ steigt die Schwierigkeit, die Handlungsträger und Kausalzusammenhänge aufzuklären.⁷⁰ Die Zurechnung dieser Vielzahl sozialer und digitaler Prozesse an einen einzelnen (individuellen oder kollektiven) Akteur stößt an ihre Grenzen. Insofern wird aus systemtheoretischer (sich von Organisationen lösender und auf die Handlungen abstellender) Sicht vorgeschlagen, als Zurechnungspunkt nicht den einzelnen Entscheidungsträger (ob Mensch, Organisation, Netzwerk oder KI-System als Handlungsträger), sondern die autonome Entscheidung selbst (als Handlung) in die Verantwortung zu nehmen.⁷¹ Dabei soll es letztlich um eine Art kollektives Risikomanagement zur Steuerung zukünftigen Verhaltens (wie die Programmierung und die Einsatzbedingungen von Algorithmen) gehen, so dass Risiken bereits präventiv eingedämmt werden.⁷² In Ergänzung hierzu kommt auch eine retrospektive Betrachtung infrage, um aus aufgetretenen Fehlern für die Zukunft lernen zu können.

⁶⁵ Bergmann, MSchrKrim 2016, 3 (10); Kölbel, NZWiSt 2018, 407 (408) m.w.N. in Fn. 14 und 15; Cornelius, ZRP 2019, 8 (10).

⁶⁶ Teubner, AcP 218 (2018), 155 (198 f.).

⁶⁷ Vgl. Gruber, in: Günther/Hilgendorf (Hrsg.), Robotik und Gesetzgebung, 2013, S. 123 (158); Klingbeil, JZ 2019, 718 (723 f.).

⁶⁸ Teubner, AcP 218 (2018), 155 (200 f.).

⁶⁹ Vgl. Beck, in: Hilgendorf/Seidel (Hrsg.), Robotics, 2017, S. 227 (233 f.); Spiecker, CR 2016, 698 (701); Cornelius, ZRP 2019, 8 (10).

⁷⁰ Vgl. Zech (Fn. 21), S. 170; Cornelius, ZRP 2019, 8 (10).

⁷¹ Teubner, AcP 218 (2018), 155 (202 f.); Cornelius, ZRP 2019, 8 (10).

⁷² Teubner, AcP 218 (2018), 155 (203); Cornelius, ZRP 2019, 8 (10).

III. Einhaltung von Rechtsregeln

Die „Hochrangige Expertengruppe“ hat vollkommen zu Recht als erste Komponente für eine vertrauenswürdige KI darauf hingewiesen, dass sicherzustellen ist, dass Entwicklung, Produktion und Vertrieb sowie die Nutzung von KI-Systemen unter Einhaltung von rechtlichen Vorgaben zu erfolgen haben.⁷³ Dabei geht es nicht allein um die implementierte Technik im KI-System selbst, sondern es muss auf eine ganzheitliche Systemgestaltung geachtet werden, was auch die Einbeziehung von organisatorischen Prozessen, vertraglichen Zusammenhängen und (bei einem Unternehmen) unternehmensinternen Vorgaben bedeutet.

1. Ausgangspunkt

Relevante Rechtsquellen sind nicht nur das Primär- (also die EU-Verträge nebst der EU-Grundrechtecharta) und das Sekundärrecht (wie die Datenschutzgrundverordnung, aber auch Richtlinien zur Antidiskriminierung, Produktsicherheit und -haftung, Verbraucher- und Gesundheitsschutz) der Europäischen Union, sondern auf internationaler Ebene auch die Menschenrechtsverträge der Vereinten Nationen und die Übereinkommen des Europarates (insbesondere die Europäische Menschenrechtskonvention) sowie die nationalen Vorschriften.⁷⁴ Dabei ist zu beachten, dass es in Abhängigkeit von dem jeweiligen Anwendungsbereich der KI-Systeme zur Anwendung ganz unterschiedlicher rechtlicher Vorschriften kommen kann (z.B. sind bei einem Medizinprodukt die entsprechenden rechtlichen Vorschriften wie die Verordnung über Medizinprodukte im Gesundheitssektor oder das Medizinproduktegesetz zu beachten).⁷⁵

2. KI-Compliance

Damit geht es im Endeffekt um eine typische Aufgabe, die der sog. Compliance zugeschrieben wird – das KI-System muss compliant sein (KI-Compliance). Die KI-Compliance selbst betrifft einerseits die Frage, ob die KI-Systeme alle relevanten allgemein geltenden rechtlichen Vorgaben nachweislich eingehalten haben.⁷⁶ Andererseits können auch solche Risiken berücksichtigt werden, die aus sonstigen externen Regelwerken (wie Verträgen, Normen oder Standards, Zertifikaten oder Richtlinien anderer Institutionen) aber auch internen Festlegungen (wie Unternehmensrichtlinien, Organisations- oder Verfahrensanweisungen, Service-Level-Agreements) resultieren.⁷⁷ Überdies zählen zu den rechtlichen Vorgaben auch die Rechtsprechung, durch die die jeweiligen Rechtsnormen und auch Sorgfaltspflichten konkretisiert und soweit notwendig auch den geänderten gesellschaftlichen Rahmenbedingungen angepasst werden (soweit dies mit dem

⁷³ Siehe oben Fn. 9.

⁷⁴ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 8 Abschnitt (21).

⁷⁵ Vgl. Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 7 Abschnitt (20).

⁷⁶ Vgl. insoweit zur IT-Compliance Klotz/Dorn, HMD 2008, Heft 5 Bd. 263, 5 (8).

⁷⁷ Klotz/Dorn, HMD 2008, Heft 5 Bd. 263, 5 (11).

strafrechtlichen Gesetzlichkeitsprinzip nach Art. 103 Abs. 2 GG vereinbar ist).⁷⁸ Hierzu können Parallelen zum methodischen Ansatz der Automatisierung von Compliance ins Auge gefasst werden. Dieser sieht vor, dass rechtliche Vorgaben und jeweils relevante Regelwerke in IT-Systeme übersetzt werden, wobei für die Übersetzung der Compliance-Anforderungen in eine formale Form auf Policy-Sprache zurückgegriffen wird.⁷⁹ Jedoch sind – aus technischer Sicht – die jeweiligen Besonderheiten von KI-Systemen zu berücksichtigen. Denn KI-Systeme folgen gerade nicht – wie logische (automatisierte) Systeme – einer strengen Logik.

3. Erklärbare KI (oder explainable AI)

Hierbei kommt die Forschungsrichtung der so genannten erklärbaren KI (oder explainable AI) ins Spiel.⁸⁰ Durch sie soll sichergestellt werden, dass die jeweiligen Prozesse und Entscheidungen der KI-Systeme nachvollziehbar und erklärbar sind.⁸¹ Dabei geht es auch darum aufzuklären, was Ursache und Wirkung und was ggf. nur eine Korrelation ist.⁸² Letztlich ist diese Erklärbarkeit⁸³ nicht nur notwendig, damit die Nutzer ein Vertrauen zu den KI-Systemen entwickeln können,⁸⁴ sondern auch für die Frage der Feststellung von Verantwortlichkeiten für den Fall, dass es zu ungewollten Ergebnissen (wie beispielsweise einer Körperverletzung von

Menschen) kommt.⁸⁵ Durch die Gewährleistung der Rückverfolgbarkeit und Nachprüfbarkeit der Daten, Algorithmen und Prozesse, die zu den entsprechenden Entscheidungen eines KI-Systems geführt haben, können nicht nur zukünftige Fehler verhindert, sondern auch die Gründe für eine fehlerhafte Entscheidung der KI ermittelt werden.⁸⁶ Eine besondere Bedeutung hat dies für sensible Situationen menschlicher Entscheidungsfindung (wie bei medizinischen Entscheidungsunterstützungssystemen), in denen den menschlichen Entscheidern zumindest eine Chance auf Überprüfung der Plausibilität der Ergebnisse eingeräumt werden muss, indem diese nachvollziehbar, transparent und erklärbar gemacht werden.⁸⁷ Dies gilt umso mehr, als aus Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g, Art. 15 Abs. 1 lit. h und Art. 22 DSGVO, insbesondere aus dem Recht auf aussagekräftige Informationen über die involvierte Logik, ein „Recht auf Erklärbarkeit“ bei automatisierten Einzelentscheidungen abgeleitet werden wird.⁸⁸

Allerdings ergibt sich eine besondere Schwierigkeit daraus, dass es nicht immer möglich ist, zu erklären, warum ein KI-System (insbesondere wenn es auf der Basis neuronaler Netze arbeitet) zu bestimmten Ergebnissen gekommen ist, also letztlich der Zusammenhang zwischen Eingabe- und Ausgabewerten nicht explizit darstellbar ist.⁸⁹ So ist es beispielsweise bei Deep-Learning-Algorithmen⁹⁰ eine besondere Herausforderung, zu erklären, warum es gerade zu bestimmten Ausgabewerten kommt und wie sich dies auf die Datenverarbeitung im gesamten Netzwerk auswirkt.⁹¹ Dennoch wäre es ein Fehlschluss, deshalb davon auszugehen, dass eine

⁷⁸ Vgl. zum an die Rechtsprechung gerichteten Optimierungs- bzw. Präzisionsgebot BVerfGE 126, 170 (198); Cornelius, GA 2015, 101 (118); ders., Verweisungsbedingte Akzessorität, 2016, S. 352 f.

⁷⁹ Ausführlich hierzu Sackmann, HMD 2008, Heft 5 Bd. 263, 39 ff.

⁸⁰ Walt/Vogl, Jusletter IT, 22.2.2018, abrufbar unter <https://www.matthes.in.tum.de/pages/1hqrk3h1m4l6/Explainable-Artificial-Intelligence-the-New-Frontier-in-Legal-Informatics> (3.2.2020).

⁸¹ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 27 Abschnitt (99).

⁸² Holzinger, in: Gesellschaft für Informatik (Hrsg.), *Informatiklexikon*, Stand: 23.4.2018, abrufbar unter <https://gi.de/informatiklexikon/explainable-ai-ex-ai/> (3.2.2020).

⁸³ Beim Verstehen als Brücke zwischen Wahrnehmen und Entscheiden kommt es auf die intellektuelle Erfassung des Zusammenhanges (den Kontext) an, während es beim Erklären (Interpretieren) darüber hinaus darum geht, die Ursachen eines beobachteten Sachverhaltes mit seinen logischen und kausalen Zusammenhängen sprachlich zu beschreiben. Dabei sind für den Menschen Bild, Sprache und Text direkt verständlich, nicht aber abstrakte Vektorräume, die dann in eine für den Menschen verständliche Form „zurückzuübersetzen“ sind, vgl. Holzinger, in: Gesellschaft für Informatik (Hrsg.), *Informatiklexikon*, Stand: 23.4.2018, abrufbar unter <https://gi.de/informatiklexikon/explainable-ai-ex-ai/> (3.2.2020).

⁸⁴ Zu diesem Aspekt vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 16 Abschnitt (53).

⁸⁵ Deshalb formuliert Herberger, NJW 2018, 2825 (2828), auch völlig zu Recht, dass „das Einfordern von Erklärungen in juristischer Hinsicht unabdingbar“ ist.

⁸⁶ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 22 Abschnitt (76), zum Grundsatz der Rückverfolgbarkeit sowie S. 24 Abschnitt (88) zur Nachprüfbarkeit.

⁸⁷ Holzinger, in: Gesellschaft für Informatik (Hrsg.), *Informatiklexikon*, Stand: 23.4.2018, abrufbar unter <https://gi.de/informatiklexikon/explainable-ai-ex-ai/> (3.2.2020); vgl. auch He/Baxter/Xu/Zhou/Zhang, *Nature Medicine* 25 (2019), 30 (32); Gang Luo, *Health Information Science and System (HISS)* Bd. 4 Nr. 2 März 2016, 1, abrufbar unter http://pages.cs.wisc.edu/~gangluo/explain_prediction_results.pdf (3.2.2020).

⁸⁸ Dix, in: Simitis/Hornung/Spiecker genannt Döhmman (Hrsg.), *Nomos Kommentar, Datenschutzrecht*, 2019, DSGVO Art. 25 Rn. 25 m.w.N., wobei die Reichweite umstritten ist, vgl. Wachter/Mittelstadt/Floridi, *Science Robotics*, 2 (6), die davon ausgehen, dass „only a general, easily understood overview of system functionality is likely to be required“.

⁸⁹ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 27 Abschnitt (99).

⁹⁰ Vgl. hierzu oben zu Fn. 13.

⁹¹ Yuan, RW 2018, 477 (490); *Sachverständigenrat für Verbraucherfragen*, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren, 2018, S. 54 f.

Evaluation dieser Systeme nicht möglich ist.⁹² Vielmehr ist es gerade bei diesen „Blackbox“-Algorithmen⁹³ notwendig, besonders auf die Rückverfolgbarkeit, Nachprüfbarkeit und transparente Kommunikation des Systems zu achten.⁹⁴ Dies bedeutet, dass die Daten, das System und die Geschäftsmodelle als systemrelevante Komponenten der KI für die Anwender transparent sein müssen.⁹⁵ Denn letztlich beruhen auch diese Algorithmen auf einem Training und damit auf einem menschlichen Verhalten. Der Mensch definiert, welche Aspekte vom System als positiv oder negativ anzusehen sind.⁹⁶

4. Präventiver Ansatz: KI-Compliance by Design durch technologische Impossibility Structures

Mittels technikbezogener Vorgaben kann präventiv darauf hingewirkt werden, dass durch die KI-Systeme erst gar kein dem geltenden Recht widersprechendes – erst recht ein für den Menschen schädliches Verhalten – ausgeführt bzw. erlernt werden kann.⁹⁷ Solch eine Begrenzung der technischen Möglichkeiten kann möglichst frühzeitig – am besten bereits in der Entwicklungsphase des Systems – durch einen „By-Design Ansatz“ berücksichtigt werden.⁹⁸ Terminologisch kann hierfür auch der Begriff der impossibility structures fruchtbar gemacht werden, der für Mechanismen steht, die Rechtsverstöße bereits rein physisch unmöglich machen.⁹⁹ Allerdings geht es hier darum, dass es bereits dem KI-System selbst unmöglich gemacht werden soll, sich außerhalb der Grenzen des Rechts (bzw. auch weitergehender Vorgaben) zu bewegen.¹⁰⁰

Ein solcher Ansatz erfordert, dass bereits ex ante die möglichen Aktivitäten und Zustände einschließlich etwaiger Ausnahmen erkennbar sind.¹⁰¹ Auch wenn dies in Reinform nicht möglich ist, da alle Aktivitäten und Zustände nicht

voraussehbar sind, lässt sich durch eine partielle Umsetzung die Komplexität verringern.¹⁰² Dabei können zwei unterschiedliche Strategien verfolgt werden: Einerseits können in Form einer „Blacklist“ diejenigen Aktivitäten und Zustände spezifiziert werden, die als nicht-compliant angesehen und deshalb nicht zugelassen werden.¹⁰³ Andererseits können in Form einer Positivliste die Aktivitäten und Zustände spezifiziert werden, die compliant sind.¹⁰⁴ Soweit dann Abweichungen von diesen Zuständen festgestellt werden, bedeutet dies noch nicht zwingend, dass ein Verstoß gegen Compliance-Vorschriften vorliegt, aber dass zusätzliche Maßnahmen im Sinne des retrospektiven Ansatzes zu ergreifen sind.¹⁰⁵

5. Retrospektiver Ansatz: KI-Compliance by detection durch intelligent surveillance

Der retrospektive Ansatz der KI-Compliance by Detection zielt darauf, dass durch eine vollständige und korrekte Aufzeichnung von tatsächlichen Aktivitäten und Zuständen eine ex-post-Bewertung ermöglicht wird, die den Abgleich der tatsächlichen Ergebnisse mit den Compliance-Anforderungen ermöglicht.¹⁰⁶ Hier können Parallelen zur Diskussion um eine technologisch durchgeführte intelligent surveillance gezogen werden. Während es durch die impossibility structures bereits objektiv ausgeschlossen wird, dass sich ein KI-System außerhalb rechtlicher oder vertraglicher Grenzen bewegt, ist ein entsprechendes Agieren bei einer intelligent surveillance noch möglich, wobei dann die aus dem Rechts- oder Vertragsbruch herrührenden Konsequenzen drohen.¹⁰⁷ Da so unerwünschte Aktivitäten oder Zustände nicht verhindert werden können, besteht die Herausforderung darin, technologische Werkzeuge zur Erkennung von Verstößen bereitzustellen, so dass einerseits gesetzte Anreize als tatsächlich erreichbar angesehen werden und andererseits auf die Verletzung von Compliance-Anforderungen (insbesondere von Rechtsverletzungen) schnell reagiert und gegebenenfalls gegengesteuert werden kann.¹⁰⁸ Neben dem nur eingeschränkt möglichen Einsatz von technischen Mechanismen wie der Unterbrechung des aktuellen Prozesses zur Herstellung eines früheren, regelkonformen Zustandes (Rollback) oder dem Starten weiterer Prozesse zur automatischen Beseitigung der für die Regelwidrigkeit verantwortlichen Bedingungen im KI-System (Recover) kommt insbesondere eine Information einer verantwortlichen Person (beispielsweise im

⁹² Yuan, RW 2018, 477 (491).

⁹³ Walzl/Vogl, Jusletter IT, 22.2.2018, S. 3, abrufbar unter <https://www.matthes.in.tum.de/pages/1hqrlk3h1m4l6/Explained-Artificial-Intelligence-the-New-Frontier-in-Legal-Informatics> (3.2.2020).

⁹⁴ Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 16 Abschnitt (53).

⁹⁵ Vgl. Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 3), S. 22 Abschnitt (75).

⁹⁶ Yuan, RW 2018, 477 (488), zum „Reinforcement Learning“, vgl. hierzu auch Fn. 26.

⁹⁷ Vgl. Hilgendorf (Fn. 21), S. 110 f.; vgl. auch Art. 6 Ziff. 2 des Verordnungsvorschlags der EU-Kommission KOM (2018) 640 endg. v. 12.9.2018, wonach Onlinedienste dazu verpflichtet werden sollen, proaktive Maßnahmen „einschließlich automatisierter Werkzeuge“ zu verwenden, um den Upload von terroristischer Kommunikation zu verhindern.

⁹⁸ Vgl. Bedner, Cloud Computing, 2012, S. 12.

⁹⁹ Rademacher, JZ 2019, 702 m.w.N. in Fn. 2.

¹⁰⁰ Zur Frage der Handlungsfähigkeit von autonom agierenden KI-Systemen bereits oben II. 3. a) Autonomierisiko.

¹⁰¹ Allgemein zum „Compliance by Design“-Ansatz Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

¹⁰² Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

¹⁰³ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

¹⁰⁴ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

¹⁰⁵ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

¹⁰⁶ Allgemein zum „Compliance by Detection“-Ansatz Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹⁰⁷ Vgl. – auf das Handeln von Menschen bezogen – Rademacher, JZ 2019, 702 (703).

¹⁰⁸ Vgl. zur automatisierten Compliance by Detection Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (43).

Unternehmen des Compliance-Officers oder eines KI-Sicherheitsbeauftragten¹⁰⁹) in Betracht.¹¹⁰

6. Kombination beider Ansätze

Der präventive Ansatz hat auf den ersten Blick den Charme, dass – in einer idealen Welt – Rechtsverstöße von vornherein unmöglich sind.¹¹¹ Jedoch stellt sich unmittelbar die Frage, ob es überhaupt möglich ist, dass die Grenze zwischen Recht und Unrecht für ein künstliches System erkennbar ist oder ob es nicht vielmehr zu einer Streubreite dahingehend kommt, dass auch Erlaubtes verhindert wird.¹¹² Wenn auch diese Gefahr des Overblockings in dem Maße geringer wird, in welchem die KI-Systeme zu einer ausreichenden Kontexterkenkung in der Lage sind¹¹³ – was bei einer so genannten starken künstlichen Intelligenz mit der (erwarteten) Fähigkeit, die Rechtmäßigkeit eines Verhaltens deduktiv aus Regeln abzuleiten prognostiziert wird¹¹⁴ –, ist es derzeit schlichtweg unmöglich, alle (!) möglichen Varianten und Zustände bereits ex ante auf ihre Vereinbarkeit mit den jeweiligen Compliance-Anforderungen bewerten zu können. Dies führt bei einer strikten Erzwingung bzw. dem unbedingten Ausschluss bestimmter Aktivitäten dazu, dass auf unvorhergesehene Ereignisse oder einen sich ändernden Kontext nicht mehr angemessen reagiert werden kann.¹¹⁵

So zeigt die Erfahrung, dass mit einem „By-design“-Ansatz auch erhebliche Nachteile einhergehen. Eine besonders konsequente Form fand sich in dem Digital Rights Management (DRM) der Musik- und Filmindustrie. Dadurch wurde versucht, bereits durch die implementierte Technik sicherzustellen, dass Urheberrechtsverletzungen durch die Nutzer ausgeschlossen wurden. Jedoch wurde das DRM auf der Seite der Kunden als eine starke Gängelung empfunden und abgelehnt.¹¹⁶ Als besonders gravierender Nachteil erwies sich die Tatsache, dass teilweise selbst die rechtmäßige Nutzung nicht mehr möglich war, wie dies durch die Abschaltung des Dienstes Playforsure von Microsoft passierte.¹¹⁷ Ebenso zeigt die Diskussion um den berühmt-berüchtigten so genannten Upload-Filter die Brisanz von impossibility structures. So sollte den Plattformbetreibern nach dem ursprünglichen Entwurf der Urheberrechts-Richtlinie explizit vorgeschrieben werden, dass diese technologische Maßnahmen ergreifen müssen, damit es von vornherein unmöglich ist, urheberrechtlich geschütztes Material ohne eine entsprechen-

de Lizenz hochzuladen.¹¹⁸ Dies führte zur Diskussion um das urheberrechtliche Overblocking, wonach ein Hochladen auch rechtmäßiger Inhalte durch die Schaffung solcher impossibility structures von vornherein verhindert wird.¹¹⁹

Deshalb ist ein alleiniger „By-design“-Ansatz durch die technologische Implementierung von impossible structures nach wie vor nicht umsetzbar. Dessen Nachteile lassen sich durch eine Kombination mit dem retrospektiven Ansatz vermeiden. Durch diesen werden die Handlungsaktivitäten nicht beschränkt und es ist keine ex-ante-Kennntnis aller möglichen Aktivitäten und Zustände notwendig.¹²⁰ Damit bietet dieser Ansatz den größtmöglichen Handlungsspielraum und die Flexibilität, auch auf unvorhergesehene Ereignisse adäquat reagieren zu können.¹²¹ Allerdings ist dies mit dem Nachteil verbunden, dass so keine präventiven Handlungsbeschränkungen möglich sind.

Bei der Wahl eines ausgewogenen Verhältnisses beider Ansätze zueinander gilt es, die jeweiligen Vor- und Nachteile sachgemäß gegeneinander abzuwägen.¹²² Das ist die Sache der im jeweiligen Lebenszyklus¹²³ des KI-Systems Verantwortlichen.

7. Regelungsvorschlag in Anlehnung an Art. 25 Abs. 1 DSGVO

Der Gesetzgeber selbst könnte jedoch entsprechende Rahmenbedingungen schaffen. Da die zu beachtenden rechtlichen Vorgaben ständigen Änderungen unterworfen sind, die sich nicht allein in einem Eingreifen des Gesetzgebers äußern, sondern auch in der Dynamik der Rechtsprechung, entsteht ein ständiger Anpassungsdruck,¹²⁴ der dabei zu berücksichtigen wäre. Die DSGVO sieht bereits eine Regelung zum „Privacy by Design“ bzw. „Datenschutz durch Gestaltung“ in Art. 25 Abs. 1 vor.¹²⁵ Hier würde sich eine über den Bereich des Datenschutzes hinausgehende verallgemeinernde Regelung anbieten.¹²⁶ In Analogie dazu könnte eine Vorgabe für KI-Systeme wie folgt ausgestaltet sein:

„Unter Berücksichtigung des Stands der Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke des Einsatzes von KI-Systemen

¹⁰⁹ Zu diesem Vorschlag im Rahmen einer angedachten Kollektivverantwortlichkeit eines Verbandes vgl. die Ausführungen zu Fn. 181.

¹¹⁰ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹¹¹ Zur Ambivalenz einer solchen „idealen“ Welt und der Forderung nach einem „Recht zum Rechtsverstoß“ Rademacher, JZ 2019, 702.

¹¹² Rademacher, JZ 2019, 702 (706).

¹¹³ Rademacher, JZ 2019, 702 (707).

¹¹⁴ Eberl, APuZ 6-8/2018, 8 (12 f.).

¹¹⁵ Vgl. Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (45).

¹¹⁶ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹¹⁷ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹¹⁸ Zwar ist diese explizit noch in Art. 13 Abs. 1 des Vorschlags für eine Richtlinie über das Urheberrecht im digitalen Binnenmarkt (KOM [2016] 593 endg.) vorgesehene Regelung nicht mehr im Art. 17 der Richtlinie 2019/790 enthalten. Dennoch ist nicht klar, ob damit Upload-Filter ausgeschlossen sind oder nicht; hierzu Rademacher, JZ 2019, 702 (704) m.w.N. in Fn. 12.

¹¹⁹ Spindler, NJW 2017, 2305 (2307).

¹²⁰ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹²¹ Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (44).

¹²² Sackmann, HMD 2008, Heft 5 Bd. 263, 39 (45).

¹²³ Zum Lebenszyklus vgl. oben zu Fn. 10.

¹²⁴ Vgl. Rüpke/v. Lewinski/Eckhardt, Datenschutzrecht, 2018, § 20 Rn. 9 ff.

¹²⁵ Zur Differenzierung dieser Begriffe vgl. Hansen, in: Simitis/Hornung/Spiecker genannt Döhmman (Fn. 88), DSGVO Art. 25 Rn. 23.

¹²⁶ Herberger, NJW 2018, 2825 (2828).

sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere der mit der Verarbeitung verbundenen Risiken für die Rechte Dritter und der Allgemeinheit sind durch die verantwortliche Stelle sowohl zum Zeitpunkt der Festlegung der Mittel für den Einsatz des KI-Systems als auch zum Zeitpunkt des eigentlichen Einsatzes geeignete technische und organisatorische Maßnahmen vorzusehen, die dafür ausgelegt sind, die Rechte Dritter und der Allgemeinheit wirksam zu sichern.“

Dabei haben Erfahrungen mit dem „Privacy by Design“-Ansatz gezeigt, dass diesem zunächst eine echte Durchschlagskraft verwehrt blieb, da es weder eine Sanktionierung noch Anreize (beispielsweise durch eine entsprechende Zertifizierung) gab.¹²⁷ Dies hat sich unmittelbar mit Einführung der DSGVO geändert. Nach Art. 83 Abs. 4a DSGVO sind bei einer Nichteinhaltung von allgemeinen Pflichten der Verantwortlichen oder Auftragsverarbeiter Geldbußen zu verhängen. Dazu gehören auch Verstöße gegen einen dem Stand der Technik entsprechenden Datenschutz, wonach die IT-Systeme so gestaltet sein müssen, dass sie die wirksame Umsetzung von Datenschutzgrundsätzen („privacy by design“) und die Forderung nach datenschutzfreundlichen Voreinstellungen („privacy by default“) gem. Art. 25 DSGVO fördern.¹²⁸

IV. Verantwortlichkeiten beim Einsatz von KI-Systemen

Bei der Bestimmung der Verantwortlichkeiten kommt aus der Sicht der Fahrlässigkeitsdelikte eine Verschiebung der Sorgfaltsmaßstäbe in Betracht (hierzu 1.). Ferner käme es in Betracht, an eine Sanktionierung der KI-Systeme selbst zu denken (siehe hierzu 2.). So ließe sich ein „hinkendes Rechtssubjekt“ vermeiden, dem zwar die Fähigkeit zu rechtsgeschäftlichem und tatsächlichem Handeln (grundsätzlich aber im fremden Interesse) zugestanden wird, welches aber für rechtswidrige Taten nicht sanktioniert werden kann.¹²⁹ Außerdem sind die Möglichkeiten einer sanktionsrechtlichen Verbands(mit)verantwortlichkeit (hierzu 3.) sowie die Sicherstellung der menschlichen Aufsicht (hierzu 4.) zu bedenken.

1. Strafrechtliche Fahrlässigkeitshaftung

Bei der Entwicklung und dem Einsatz von KI-Systemen kann es bei den Fahrlässigkeitsdelikten (hier sind insbesondere die §§ 222, 229 StGB zu nennen) zu einer Verschiebung von einzuhaltenden Sorgfaltsmaßstäben kommen. Denn für eine Fahrlässigkeitshaftung kommt es insbesondere auf eine Sorgfaltspflichtverletzung an.¹³⁰ Dabei ist nach der überwiegenden Auffassung im Strafrecht auf der Ebene der Tatbestandsmäßigkeit ein objektiver Maßstab an die einzuhalten-

den Sorgfaltsanforderungen anzulegen, während auf der Ebene der persönlichen Schuld danach zu fragen ist, inwieweit der Täter nach seinen individuellen Fähigkeiten subjektiv in der Lage war, den objektiven Sorgfaltsanforderungen zu entsprechen.¹³¹ Hier kommt es entscheidend auf den Umfang von Verhaltensregeln an, die den Sorgfaltsmaßstab bilden und sich aus Rechtssätzen, aber auch aus nichtstaatlichen Regeln (wie Industrienormen) ergeben können.¹³² Der Gesetzgeber selbst greift dabei auf generalklauselartige Beschreibungen wie den „Stand der Wissenschaft und Technik“ in § 1 Abs. 2 Nr. 5 ProdHG oder den „Stand der medizinischen Wissenschaft und Technik“ in § 5 Abs. 1 TFG zurück.¹³³ Soweit solche speziellen Verhaltensregeln nicht existieren, wird auf den „besonnenen und gewissenhaften Menschen aus dem Verkehrskreis des Täters“ abgestellt.¹³⁴

Damit ist – soweit es aufgrund eines Programmierfehlers zur Schädigung (beispielsweise der körperlichen Integrität eines Menschen) durch ein KI-System kommt – die Frage zu beantworten, wie sich ein sorgfältiger Programmierer (bzw. KI-System-Entwickler) verhalten hätte bzw. ob ggf. eine Übernahmefahrlässigkeit vorgeworfen werden kann. Wegen der Komplexität bietet sich bei der Herstellung (bzw. Programmierung oder dem Training) von KI-Systemen eine Vielzahl von Anknüpfungspunkten für menschliche Sorgfaltspflichtverletzungen an.¹³⁵ Darüber hinaus müssen die Hersteller auch während des Betriebes der KI-Systeme noch produktbezogene Verkehrssicherungspflichten beachten.¹³⁶

Es ist vorstellbar, dass der *Nichteinsatz* verfügbarer Systeme als eine Sorgfaltspflichtverletzung angesehen wird, wenn es dadurch zu vermeidbaren Schädigungen kommt.¹³⁷ So liegt im medizinischen Bereich ein Behandlungsfehler

¹³¹ Sternberg-Lieben/Schuster, in: Schönke/Schröder, Strafgesetzbuch, Kommentar, 30. Aufl. 2019, § 15 Rn. 118.

¹³² Sternberg-Lieben/Schuster (Fn. 131), § 15 Rn. 135; jedoch ist hierbei zu beachten, dass vor dem Hintergrund der schadensorientierten Haftungsprinzipien des Zivilrechts keine unbesehene Übernahme zivilrechtlicher Sorgfaltsanforderungen für die Frage der strafrechtlichen Verantwortlichkeit erfolgen darf, BGH NJW 1990, 2560 (2562); ausführlich zu dieser Problematik Blechschmitt, Die straf- und zivilrechtliche Haftung des Arztes beim Einsatz roboterassistierter Chirurgie, 2017, S. 129 ff., mit der Nennung des Eckpunktes (S. 138), dass ein zivilrechtlicher Maßstab die Obergrenze möglicher Fahrlässigkeitsstrafbarkeit bildet, aber noch nicht zwingend zu einem strafrechtlich relevanten Sorgfaltsmangel führen muss.

¹³³ Yuan, RW 2018, 477 (494 f.).

¹³⁴ Rengier, Strafrecht, Allgemeiner Teil, 11. Aufl. 2019, § 52 Rn. 15; als „völlig nichtssagende Kunstfigur“ kritisierend Duttge, in: Joecks/Miebach (Hrsg.), Münchener Kommentar zum Strafgesetzbuch, Bd. 1, 3. Aufl. 2017, § 15 Rn. 117.

¹³⁵ Yuan, RW 2018, 477 (496), mit Beispielen.

¹³⁶ Yuan, RW 2018, 477 (496 Fn. 95 m.w.N.).

¹³⁷ So bereits Kilian, in: Savory (Hrsg.), Expertensysteme: Nutzen für Ihr Unternehmen, 2. Aufl. 1989, S. 393 (395), mit Blick auf eine nicht erfolgte Konsultation verfügbarer medizinischer Expertensysteme.

¹²⁷ Hansen (Fn. 125), DSGVO Art. 25 Rn. 6.

¹²⁸ Cornelius, in: Forgó/Helfrich/Schneider (Hrsg.), Betrieblicher Datenschutz, 3. Aufl. 2019, Teil XIV Rn. 61.

¹²⁹ Vgl. Wagner, NZWiSt 2018, 399 (400), allerdings im Hinblick auf die Verbandsstrafbarkeit.

¹³⁰ Kühl, Strafrecht, Allgemeiner Teil, 8. Aufl., 2017, § 17 Rn. 14 ff.

dann vor, wenn eine gewählte Behandlungsmethode „veraltet und überholt“ ist, was dann anzunehmen ist, „wenn neue Methoden risikoärmer sind und/oder bessere Heilungschancen versprechen, in der medizinischen Wissenschaft im Wesentlichen unumstritten sind und deshalb nur ihre Anwendung von einem sorgfältigen und auf Weiterbildung bedachten Arzt verantwortet werden kann“.¹³⁸ Dies dürfte beispielsweise dann der Fall sein, wenn KI-Systeme so leistungsfähig sein sollten, dass deren Diagnose der Vorzug vor einer abweichenden ärztlichen Meinung gegeben wird, der Einsatz eines solchen Systems also medizinischer Standard ist.¹³⁹ Das ist keine in der Ferne spielende Musik, sondern für bestimmte medizinische Konstellationen aktuell. So ist bereits heute der Nachweis gelungen, dass deep-learning-Systeme besser als Dermatologen in der Lage sind, Hautkrebs zu erkennen.¹⁴⁰

Wenn dagegen ein solches KI-System eingesetzt wird und sich die Diagnose im Nachhinein als fehlerhaft herausstellt, kommt eine Sorgfaltspflichtverletzung der (ärztlichen) Anwender nur bei einer fehlerhaften Auswahl (keine Kompatibilität des Systems mit dem Einsatzzweck), Bedienung (nicht entsprechend den Herstellervorgaben) oder Überwachung (keine regelmäßige Wartung, keine Beachtung offensichtlicher Sicherheitsmängel) des Systems in Betracht.¹⁴¹ Dabei ist zu beachten, dass sich die jeweiligen Sorgfaltsanforderungen gegenseitig beeinflussen. So kann eine besonders sorgfältige Auswahl eines KI-Systems zu einer Verringerung der Überwachungspflichten führen (und andersherum).¹⁴²

Daneben kommt noch eine Herstellerhaftung in Betracht.¹⁴³ Für diese ist es jedoch Voraussetzung, dass es sich um einen vermeidbaren und behebbaren Konstruktions- oder Trainingsfehler und nicht um eine Fehlreaktion im Einzelfall handelt.¹⁴⁴ Dabei kann es zu einer Verschiebung der Sorgfaltspflichten weg vom Nutzer hin zum Hersteller kommen.¹⁴⁵

Weitere Voraussetzung für eine strafrechtliche Fahrlässigkeitshaftung ist die Vorhersehbarkeit des Erfolgesintritts im Zeitpunkt der Sorgfaltspflichtverletzung. Die Rechtsprechung stellt hauptsächlich auf den Eintritt des tatbestandlichen Erfolges ab,¹⁴⁶ während die Literatur noch die konkrete Art und Weise der Herbeiführung des Taterfolges mit einbe-

zieht.¹⁴⁷ Bei Zugrundelegung des Maßstabes der Rechtsprechung kann die Vorhersehbarkeit regelmäßig dann bejaht werden, wenn es sich um KI-Systeme handelt, die bestimmungsgemäß mit Menschen interagieren und eine Verletzung derselben denkbar ist.¹⁴⁸

Schließlich muss der Erfolg – nach Feststellung einer Kausalität im Sinne der *conditio-sine-qua-non*-Formel¹⁴⁹ – dem Täter zurechenbar sein. Dies erfordert die Schaffung eines rechtlich missbilligten Risikos, welches sich gerade im tatbestandlichen Erfolg verwirklicht.¹⁵⁰ Dabei wird neben dem Erfordernis eines bestehenden Schutzzweckzusammenhangs¹⁵¹ auf einen Pflichtwidrigkeitszusammenhang¹⁵² abgestellt. Dieser spezifische Ursachenzusammenhang ist dann zu bejahen, wenn auf der Grundlage eines hypothetischen Geschehensverlaufs bei Hinwegdenken des vorwerfbaren Tatumstandes und Hinzudenken eines pflichtgemäßen Alternativverhaltens des Täters der tatbestandliche Erfolg mit an Sicherheit grenzender Wahrscheinlichkeit¹⁵³ entfallen würde.¹⁵⁴ Hier wird wieder deutlich, wie wichtig eine entsprechende Nachvollziehbarkeit der Entscheidungsfindung eines KI-Systems für die Bestimmung des etwaigen Eintretens aufgrund eines rechtmäßigen Alternativverhaltens ist.

2. Sanktionierbarkeit von KI-Systemen?

Neben einer Fahrlässigkeitsverantwortlichkeit der beteiligten Personen stellt sich die Frage einer Sanktionierbarkeit der KI-Systeme selbst. Dem könnte das Erfordernis der Schuld entgegenstehen, welches nach dem üblichen Verständnis Voraussetzung für die Kriminalstrafe ist.¹⁵⁵ Denn eine auf den

¹³⁸ BGH NJW 1988, 763 (764); vgl. zur Verfehlung des medizinischen Standards aufgrund Anwendung einer veralteten Behandlungsmethode *Blechschnitt* (Fn. 132), S. 121 ff.

¹³⁹ Meyer, ZRP 2018, 233 (236).

¹⁴⁰ Brinker/Hekler/Enk/Klode/Hauschild/Berking/Schilling/Haferkamp/Schadendorf/Holland-Letz/Utikal/v. Kalle, *European Journal of Cancer* 113 (2019), 47.

¹⁴¹ Horner/Kaulartz, CR 2016, 7 (8), haben diese Kategorisierung zwar unter Rückgriff auf § 831 BGB entwickelt, wobei jedoch nichts dagegenspricht, für den objektiven Maßstab auf diese Kategorisierung auch aus strafrechtlicher Sicht zurückzugreifen, vgl. oben Fn. 132.

¹⁴² Horner/Kaulartz, CR 2016, 7 (8).

¹⁴³ Meyer, ZRP 2018, 233 (236).

¹⁴⁴ Meyer, ZRP 2018, 233 (236).

¹⁴⁵ Horner/Kaulartz, CR 2016, 7 (8).

¹⁴⁶ RGSt 54, 351; BGHSt 12, 75.

¹⁴⁷ Sternberg-Lieben/Schuster (Fn. 131), § 15 Rn. 180.

¹⁴⁸ Yuan, RW 2018, 477 (497).

¹⁴⁹ BGHSt 1, 332 (333).

¹⁵⁰ Lackner/Kühl, *Strafgesetzbuch, Kommentar*, 29. Aufl. 2018, Vor § 13 Rn. 14 m.w.N.

¹⁵¹ Dieser ist dann gegeben, wenn sich der eingetretene Erfolg als Verwirklichung des Risikos erweist, vor dem die übertretene Sorgfaltspflicht schützen sollte.

¹⁵² Duttge (Fn. 134), § 15 Rn. 165.

¹⁵³ BGHSt 11, 1 (3); vgl. aber BGHSt 49, 1 (4), wonach der „Ursachenzusammenhang“ schon dann entfalle, wenn sich derselbe Erfolgesintritt bei verkehrsgerechtem Verhalten „aufgrund erheblicher Tatsachen nach der Überzeugung des Tatrichters nicht ausschließen“ lasse.

¹⁵⁴ Das Beweismaß ist umstritten. So wird in der Literatur schon das „überwiegend wahrscheinliche“ Ausbleiben des Erfolges als ausreichend angesehen; Duttge (Fn. 134), § 15 Rn. 184.

¹⁵⁵ BVerfGE 45, 187 (253 ff.); Fischer, *Strafgesetzbuch mit Nebengesetzen, Kommentar*, 67. Aufl. 2020, § 20 Rn. 2, wobei zwischen dem verfassungsrechtlichen und dem strafrechtsdogmatischen Schuldbegriff zu differenzieren ist; Hörnle, in: Sieber/Dannecker/Kindhäuser/Vogel/Walter (Hrsg.), *Strafrecht und Wirtschaftsstrafrecht, Dogmatik, Rechtsvergleich, Rechtstatsachen*, Festschrift für Klaus Tiedemann zum 70. Geburtstag, 2008, S. 325 (340); dies., *Kriminalstrafe ohne Schuldvorwurf*, S. 60, wobei sie (S. 49 ff.) unter Bezugnahme auf die Erkenntnisse der Hirnforschung

Menschenwürdegehalt des Grundgesetzes zurückgeführte Schuld¹⁵⁶ kann es für „künstliche“ Systeme nicht geben, so intelligent sie auch sein mögen.¹⁵⁷ Die Menschenwürde kann nur Menschen und nicht Tieren¹⁵⁸, Robotern und anderen „künstlichen Personen“ zukommen. Da das deutsche Strafrecht bisher nur die Strafbarkeit „natürlicher“ Personen kennt, weil (so zumindest die bisher herrschende Meinung) eine Verbandsperson weder handlungs- noch schuldfähig sei,¹⁵⁹ könnte die Betrachtung an dieser Stelle abgeschlossen werden.¹⁶⁰

Damit weicht Deutschland aber zum Teil erheblich von den Modellen anderer Staaten ab. So wird in common law-Ländern wie England, Schottland und Zypern nicht zwischen der Strafbarkeit juristischer und natürlicher Personen differenziert. Die USA kennen die strafrechtliche Verantwortlichkeit von Kapitalgesellschaften ebenso wie Spanien, Frankreich, die Niederlande und Portugal.¹⁶¹ Somit ist in den Ländern, die die Strafbarkeit juristischer Personen anerkennen, auch die Möglichkeit eröffnet, eine entsprechende strafrechtliche Haftung weiterer nicht-menschlicher Akteure zu etablieren. Hier reiht sich die aktuelle Diskussion in Deutschland um die Etablierung eines Unternehmenssanktionenrechts ein!

für die Aufgabe des Schuldbegriffs eintritt; vgl. auch (mit Blick auf juristische Personen) die Idee einer schuldgelösten Präventionsstrafe bei Wohlers, NZWiSt 2018, 412 (419), und einer parastrafrechtlichen Regelung bei Jahn/Schmitt-Leonardy/Schoop, wistra 2018, 27 (29).

¹⁵⁶ Vgl. BVerfGE 123, 267 (Lissabon), Rn. 364: „Das Strafrecht beruht auf dem Schuldgrundsatz. Dieser setzt die Eigenverantwortung des Menschen voraus, der sein Handeln selbst bestimmt und sich kraft seiner Willensfreiheit zwischen Recht und Unrecht entscheiden kann. Dem Schutz der Menschenwürde liegt die Vorstellung vom Menschen als einem geistig-sittlichen Wesen zugrunde, das darauf angelegt ist, in Freiheit sich selbst zu bestimmen und sich zu entfalten [...]. Auf dem Gebiet der Strafrechtspflege bestimmt Art. 1 Abs. 1 GG die Auffassung vom Wesen der Strafe und das Verhältnis von Schuld und Sühne [...]. Der Grundsatz, dass jede Strafe Schuld voraussetzt, hat seine Grundlage damit in der Menschenwürdegarantie des Art. 1 Abs. 1 GG [...]. Das Schuldprinzip gehört zu der wegen Art. 79 Abs. 3 GG unverfügbaren Verfassungsidentität, die auch vor Eingriffen durch die supranational ausgeübte öffentliche Gewalt geschützt ist.“

¹⁵⁷ Cornelius, ZRP 2019, 8 (10).

¹⁵⁸ Zur Diskussion im Hinblick auf Tiere als „Personen im Recht“ vgl. Raspé, Die Anerkennung der tierlichen Person im Recht, 2013, passim.

¹⁵⁹ Vgl. zu dieser Frage Jescheck/Weigend, Strafrecht, Allgemeiner Teil, 5. Aufl. 1996, § 23 VII. 1.; Lackner/Kühl (Fn. 150), § 14 Rn 1a; Laue, JURA 2010, 339; Roxin, Strafrecht, Allgemeiner Teil, Bd. 1, 4. Aufl. 2006, § 8 Rn. 58 ff.; BGHSt 3, 130; 12, 295.

¹⁶⁰ Wenn nicht auf die Voraussetzung einer Schuld an sich verzichtet wird, vgl. hierzu oben Fn. 155.

¹⁶¹ Joecks, in: Joecks/Miebach, (Fn. 134), Vor § 25 Rn. 16.

Wenn der im Koalitionsvertrag geäußerte politische Wille¹⁶² tatsächlich verwirklicht wird und eine direkte Sanktionierung von Unternehmen auch in Deutschland etabliert wird, können die angesprochenen Grundsatzfragen einer etwaigen Sanktionierbarkeit nicht-menschlicher Akteure wieder auf den Prüfstand kommen.

Das Grundgesetz selbst ist gegenüber einer Anpassung des Schuldbegriffs offen.¹⁶³ Dagegen lässt sich auch nicht die Lissabon-Entscheidung des Bundesverfassungsgerichts¹⁶⁴ anführen, denn diese bezieht sich nur auf den individualisierten – auf den Menschen bezogenen – Schuldgrundsatz.¹⁶⁵ Beispielsweise wird mit Blick auf die Diskussion eines Strafrechts von Verbänden – bei Akzeptanz des Ausgangspunktes, den Verband selbst als Akteur zu sehen – der „soziale Schuldvorwurf“ diskutiert, der auf ein Auswahl- und Organisationsverschulden des Verbandes abstellt.¹⁶⁶ Sowenig der Menschenwürdegehalt des (individualrechtlichen) Schuldprinzips der Einführung einer Verbandssanktionierung entgegensteht, sowenig steht dieser einer Sanktionierung von KI-Systemen entgegen.¹⁶⁷

Allerdings sind (auch für die Frage der Sanktionsfähigkeit) die verfolgten Strafzwecke zu berücksichtigen. Die Überlegungen zu den Verbandssanktionen stellen vorwiegend auf Prävention ab,¹⁶⁸ wobei teilweise auch der Vergeltungs-

¹⁶² Der Koalitionsvertrag zwischen CDU/CSU und SPD bekennt sich unter dem Stichwort Unternehmenssanktionen ausdrücklich hierzu: „Wir wollen sicherstellen, dass Wirtschaftskriminalität wirksam verfolgt und angemessen geahndet wird. Deshalb regeln wir das Sanktionsrecht für Unternehmen neu. Wir werden sicherstellen, dass bei Wirtschaftskriminalität grundsätzlich auch die von Fehlverhalten von Mitarbeiterinnen und Mitarbeitern profitierenden Unternehmen stärker sanktioniert werden. Bislang liegt es im Ermessen der zuständigen Behörde, ob auch das betreffende Unternehmen verfolgt wird. Durch die Abkehr vom Opportunitätsprinzip des bislang einschlägigen Ordnungswidrigkeitenrechts sorgen wir für eine bundesweit einheitliche Rechtsanwendung. Durch klare Verfahrensregelungen erhöhen wir zudem die Rechtssicherheit der betroffenen Unternehmen.“ Vgl. hierzu den bisher nur in Fachkreisen zirkulierenden Referentenentwurf des Bundesministeriums der Justiz und für Verbraucherschutz für ein Gesetz zur Bekämpfung der Unternehmenskriminalität (Bearbeitungsstand: 15. August 2019).

¹⁶³ Jahn, in: Jahn/Schmitt-Leonardy/Schoop, Das Unternehmensstrafrecht und seine Alternativen, 2016, S. 53, 73 f.; Simmler/Markwalder, ZStW 2017, 20 (41), sehen „Schuld als Zuschreibung“; vgl. auch die Begründung zum Kölner Entwurf des Verbandssanktionengesetzes, S. 22.

¹⁶⁴ Vgl. oben Fn. 156.

¹⁶⁵ Jahn (Fn. 163), S. 53 (73 f.); vgl. auch die Begründung zum Kölner Entwurf des Verbandssanktionengesetzes, S. 22.

¹⁶⁶ Tiedemann, NJW 1988, 1169 (1172 f.); Dannecker, GA 2001, 101 (112); Hilf, NZWiSt 2016, 189 (190 ff.).

¹⁶⁷ Cornelius, ZRP 2019, 8 (10).

¹⁶⁸ Kubiciel, ZRP 2014, 133 (134 ff.); Wagner, NZWiSt 2018, 399 (401 f.); Weigend/Hoven, ZRP 2018, 30 (31).

gedanke berücksichtigt wird.¹⁶⁹ Dies lässt sich zwanglos auf KI-Systeme – ebenfalls als künstliche Gebilde – übertragen. Eine verhängte Strafe könnte als „kommunikative Vergeltung“ eine vom Rechtsbruch eines KI-Systems ausgehende Botschaft im Sinne einer ausgleichenden Gerechtigkeit korrigieren.¹⁷⁰ Beim Abstellen auf den Gesichtspunkt der Generalprävention wäre eine Sanktionsfähigkeit zu bejahen, wenn die Wirkung einer verhängten Sanktion gegenüber anderen autonomen Systemen vergleichbar mit derjenigen ist, die die verhängte Sanktion gegenüber dem betroffenen System selbst hat.¹⁷¹

Dafür ist es notwendig, dass entsprechende Mechanismen etabliert werden, die die Verbreitung von rechtskräftig verhängten Sanktionen gewährleisten; sodass sie bei der Abarbeitung des Algorithmus (automatisch) mit berücksichtigt werden.¹⁷² Außerdem müsste unter dem Gesichtspunkt der Generalprävention insbesondere bei dem durch eine Sanktion betroffenen KI-System eine Funktionalität implementiert sein, die dieses davon abhält, das bestrafte Verhalten noch einmal zu begehen.¹⁷³ Dies sind notwendige Voraussetzungen, damit die Adressierung von Sanktionen direkt an KI-Systeme sinnvoll ist. Solange diese nicht gegeben sind, ist eine Sanktionierung der KI-Systeme selbst nicht zielführend.

3. Möglichkeiten einer sanktionsrechtlichen Verbands(mit-)verantwortlichkeit

Damit schließt sich die unmittelbare Frage an, inwieweit die Risiken beim Einsatz von KI (Autonomie-, Verbund- und Vernetzungsrisiko) beim kommenden Verbandssanktionenrecht Berücksichtigung finden können. Zu dieser Frage wurde bereits an anderer Stelle herausgearbeitet,¹⁷⁴ dass sich frappierende Parallelen zwischen der Frage der Steuerung etwaigen rechtswidrigen Agierens von KI-Systemen durch das Recht und der Möglichkeit der Einführung einer Verbandsverantwortlichkeit ergeben, was Anlass dafür ist, die Berücksichtigung von rechtswidrigen „Handlungen“ der KI-Systeme in eine strafrechtliche Unternehmenshaftung mit zu bedenken. So könnte eine Integration dergestalt erfolgen, dass der Normadressat von Verbandssanktionen nicht rechtlich formal, sondern funktional – an den faktischen Gegebenheiten ausgerichtet – erfolgt und sich so das (über die „Grenzen“ einer juristischen Person hinausgehende) Verbund- und insbesondere Vernetzungsrisiko bei dem Einsatz autonomer

KI-Systeme besser erfassen ließe.¹⁷⁵ Allerdings sind zuvor Kriterien zu erstellen, wie das soziale System als Normadressat selbst bestimmt werden soll, damit Rechtssicherheit (auch in der Wahrnehmung der Verteidigungsrechte) gewährleistet werden kann.

Ferner sollte eine Verantwortlichkeit des Verbandes für rechtswidrige (aber schuldlose) Handlungen von im Interesse des Verbandes eingesetzten KI-Systemen in das (noch zu schaffende) Verbandssanktionengesetz integriert werden. Der Grund für eine solche Zurechnung kann darin gesehen werden, dass dem Unternehmen (sozusagen als Prinzipal) ein rechtswidriges Verhalten des Algorithmus zugerechnet wird, welchen es zum eigenen Nutzen eingesetzt hat. Dabei käme es auf die tatsächliche Repräsentation des Verbandes (Prinzipals) durch das KI-System an, also wie sich der Repräsentant für Außenstehende darstellt.¹⁷⁶

Insofern besteht beispielsweise nach dem Kölner Entwurf des Verbandssanktionengesetzes (nachfolgend: Kölner Entwurf) durchaus die Möglichkeit dazu, da nach dessen § 1 Abs. 3 eine Zuwiderhandlung bereits bei rechtswidrigen Handlungen gegeben sein sollte, die den objektiven Tatbestand eines Strafgesetzes erfüllte. Dies ließe sich beispielsweise durch eine Ergänzung des § 1 des Kölner Entwurfs dahingehend lösen, dass zu Mitarbeitern auch autonome KI-Systeme zählen, die im Interesse des Verbandes eingesetzt werden (wem diese Gleichstellung mit menschlichen Personen missfällt, könnte auch die Kategorie der autonomen KI-Systeme separat einführen). Außerdem ist es nach dem Kölner Entwurf noch erforderlich, dass eine Organisationspflichtverletzung der Leitungsebene vorliegt.¹⁷⁷ Es ist also eine personalisierbare Zuwiderhandlung notwendig, die mit einem Leitungsversagen zusammenfällt.¹⁷⁸ Es müssen „erforderliche und zumutbare Maßnahmen zur Verhinderung“ der Zuwiderhandlungen von KI-Systemen durch eine Leitungsperson des Verbandes unterlassen worden sein.¹⁷⁹ Hier könnte dadurch Abhilfe geschaffen werden, dass von Unternehmen, die KI-Systeme einsetzen möchten, zwingend KI-Sicherheitsbeauftragte zu bestellen sind,¹⁸⁰ welche in § 1 Abs. 4 des Kölner Entwurfs als Leitungspersonen mit aufzunehmen wären.¹⁸¹ Diese hätten dann mit Blick auf den Betrieb von KI-Systemen sicherzustellen, dass sich diese tatsächlich innerhalb der durch die Gesellschaft gegebenen und in das Recht gegossenen Grenzen bewegen.¹⁸²

Inzwischen liegt ein Referentenentwurf des Bundesministeriums der Justiz und für Verbraucherschutz (mit Bearbeitungsstand vom 15. August 2019) zur Sanktionierung von verbandsbezogenen Straftaten vor (VerSanG-E). Anknüp-

¹⁶⁹ Dannecker/Dannecker, NZWiSt 129 (2016), 162 (173); v. Hirsch, NZWiSt 2016, 161, stellt im Sinne der „Tadeltheorie“ auf die normative Urteilsfähigkeit des führenden Personals ab; vgl. Jahn/Schmitt-Leonardy/Schoop, wistra 2018, 27 (28 f.).

¹⁷⁰ Cornelius, ZRP 2019, 8 (11); vgl. Beale, ZStW 126 (2014), 27 (39).

¹⁷¹ Cornelius, ZRP 2019, 8 (11), unter Hinweis auf die Argumentation zur Sanktionsfähigkeit von Verbänden bei Dannecker, GA 2001, 101 (114).

¹⁷² Cornelius, ZRP 2019, 8 (11).

¹⁷³ Cornelius, ZRP 2019, 8 (11).

¹⁷⁴ Cornelius, ZRP 2019, 8 (11 ff.).

¹⁷⁵ Cornelius, ZRP 2019, 8 (11 f.).

¹⁷⁶ Wagner, NZWiSt 2018, 399 (403).

¹⁷⁷ Kritisch dazu Wagner, NZWiSt 2018, 399 (404).

¹⁷⁸ Köbel, NZWiSt 2018, 407 (408).

¹⁷⁹ Vgl. § 3 Abs. 2 des Kölner Entwurfs.

¹⁸⁰ Vgl. insoweit die nach dem IT-Sicherheitsgesetz bisher für die Betreiber kritischer Infrastrukturen vorgesehenen IT-Sicherheitsbeauftragten.

¹⁸¹ Cornelius, ZRP 2019, 8 (12).

¹⁸² Cornelius, ZRP 2019, 8 (12).

fungspunkt für eine Verbandsverantwortlichkeit ist eine Straftat, durch die Pflichten, die den Verband treffen, verletzt worden sind oder durch die der Verband bereichert worden ist oder werden sollte (§ 2 Abs. 1 Nr. 3 VerSanG-E). Daraus wird bereits deutlich, dass es sich um eine Straftat handeln muss, die tatbestandsmäßig, rechtswidrig und schuldhaft begangen wurde.¹⁸³ Insoweit ist nach diesem Entwurf ein rechtswidriges (aber nicht schuldhaftes) „Fehlverhalten“ von KI-Systemen nicht integrierbar. Zwar ist es für eine Zurechnung einer Verbandsstraftat schon ausreichend, dass eine Leitungsperson Vorkehrungen gegen die Begehung einer solchen Tat objektiv pflichtwidrig unterlassen hat und dadurch eine objektiv erkennbare Gefahr für die Begehung einer Straftat geschaffen wurde (vgl. § 3 Abs. 1 Nr. 2 VerSanG-E). Dies ändert jedoch nichts daran, dass der Anknüpfungspunkt für eine Verbandsverantwortlichkeit eine volldektisch begangene Verbandsstraftat ist.¹⁸⁴ Insoweit sollte im Laufe des Gesetzgebungsverfahrens eine Umorientierung und Öffnung (beispielsweise im Sinne des Kölner Entwurfs) erfolgen, um entsprechendes „Fehlverhalten“ von durch Verbände betriebenen KI-Systemen, welches durch eine unzureichende Überwachung seitens von KI-Sicherheitsverantwortlichen ermöglicht wurde, gegenüber dem Verband sanktionieren zu können.

4. Menschliche Aufsicht

Insoweit kommt der Sicherstellung der menschlichen Aufsicht eine besondere Rolle zu. Eine solche Aufsicht kann durch entsprechende Lenkungs- und Kontrollmechanismen wie die interaktive Einbindung eines Menschen (Human in the Loop – HITL), die Überprüfung und Kontrolle durch einen Menschen (Human on the Loop – HOTL) oder mittels der Gesamtsteuerung durch einen Menschen (Human in Command – HIC) erreicht werden.¹⁸⁵

Der „Human in the Loop“ (HITL) mit der Fähigkeit, in jeden Entscheidungsprozess eines Systems einzugreifen, stellt die Effektivität der KI-Systeme für sich genommen in Frage, sodass dies generell nicht unbedingt angestrebt werden sollte.¹⁸⁶ Dabei ist jedoch danach zu differenzieren, um welches Einsatzgebiet der KI es sich handelt.¹⁸⁷ Wenn es beispielsweise um den Einsatz im medizinischen Bereich mit unmittelbaren Auswirkungen auf Leib oder Leben der Patien-

ten geht, wird nach wie vor der HITL erforderlich sein.¹⁸⁸ Die Verknüpfung zwischen künstlicher Intelligenz und (einem oder mehreren) Menschen zu einem „hybriden Multi-Agenten-Interaktionsmodell“ wird für diesen Bereich „als große Chance für die Zukunft“ gesehen.¹⁸⁹ Dieses Beispiel verdeutlicht einmal mehr, dass generelle Aussagen zur künstlichen Intelligenz schwierig zu treffen sind, sondern es vielmehr darauf ankommt, nach den verschiedenen Einsatzgebieten zu differenzieren.

Jedoch ist es wichtig, dass zumindest die Fähigkeit des Menschen aufrechterhalten bleibt, das Entwurfsstadium eines Systems zu steuern und den Systembetrieb zu überwachen (HOTL).¹⁹⁰ Grundsätzlich sollte der Gesamtbetrieb eines KI-Systems durch den „Human in Command“ beaufsichtigt werden und dieser Mensch die Fähigkeit haben, jederzeit über ein Ob, Wie und Wann eines Einsatzes des Systems zu entscheiden, einschließlich der Frage, inwieweit ein bestimmtes Maß an menschlichem Ermessen zugelassen und eine Entscheidung des Systems gegebenenfalls neutralisiert wird.¹⁹¹ Gerade wegen der schwierigen Vorhersehbarkeit von Entscheidungen der KI-Systeme kommt deren Überwachung eine besondere Bedeutung zu, um sicherzustellen, dass sich das System auch nach seiner Einführung wie vorgesehen verhält bzw. sich innerhalb zuvor definierter Zielsetzungen bewegt (wozu auch die Einhaltung rechtlicher Vorgaben gehört).¹⁹² Das korrespondiert auch mit dem Vorschlag, bei dem Einsatz von KI-Systemen durch Unternehmen eine entsprechend verantwortliche Person zu bestimmen, so dass etwaige rechtswidrige Aktionen der KI-Systeme bei einer fehlerhaften Überwachung durch die hierzu bestimmte Leitungsperson (wie einem KI-Sicherheitsbeauftragten)¹⁹³ direkt dem Verband zugerechnet werden können.

V. Ergebnis

Generelle Aussagen zur künstlichen Intelligenz sind mit Vorsicht zu genießen. Es ist vielmehr wichtig, auch bezüglich etwaiger regulatorischer Anforderungen nach den verschiedenen Einsatzgebieten zu differenzieren. Insoweit könnte es ein Ansatzpunkt sein, auf europäischer und nationaler Ebene eine im Bereich der KI tätige Behörde zu bestimmen, die Nutzen und Risiken von autonomen Systemen bewertet und

¹⁸³ Referentenentwurf des Bundesministeriums der Justiz und für Verbraucherschutz, Entwurf eines Gesetzes zur Bekämpfung der Unternehmenskriminalität (Bearbeitungsstand: 15.8.2019), S. 74.

¹⁸⁴ Referentenentwurf des Bundesministeriums der Justiz und für Verbraucherschutz, Entwurf eines Gesetzes zur Bekämpfung der Unternehmenskriminalität (Bearbeitungsstand: 15.8.2019), S. 78.

¹⁸⁵ *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 19 Abschnitt (65).

¹⁸⁶ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 19 Abschnitt (65).

¹⁸⁷ Meyer, ZRP 2018, 233 (235).

¹⁸⁸ Kilian (Fn. 137), S. 404; vgl. He/Baxter/Xu/Zhou/Zhang, *Nature Medicine* 25 (2019), 30, Brinker/Hekler/Enk/Klode/Hauschild/Berking/Schilling/Haferkamp/Schadendorf/Holland-Letz/Utikal/v. Kalle, *European Journal of Cancer* 113 (2019), 47 (52).

¹⁸⁹ Holzinger, in: Gesellschaft für Informatik (Hrsg.), *Informatiklexikon*, Stand: 23.4.2018, abrufbar unter <https://gi.de/informatiklexikon/explainable-ai-ex-ai/> (3.2.2020).

¹⁹⁰ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 19 Abschnitt (65).

¹⁹¹ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 19 Abschnitt (65).

¹⁹² Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz* (Fn. 3), S. 27 Abschnitt (100 f.).

¹⁹³ Vgl. oben zu Fn. 180.

entsprechende Standards diskutiert und etabliert.¹⁹⁴ Diesbezüglich könnte es gegebenenfalls angebracht sein, eine solche Aufgabe der Agentur der Europäischen Union für Cybersicherheit sowie dem Bundesamt für Sicherheit in der Informationstechnik zuzuordnen.

Um die Rechtmäßigkeit der KI-Systeme sicherstellen zu können, sollte bereits bei deren Entwicklung, aber auch bei der Einführung und Nutzung soweit wie möglich sichergestellt werden, dass die Rechtsregeln eingehalten werden (präventiver Ansatz). Außerdem wäre durch eine entsprechende Dokumentation zu ermöglichen, dass im Nachhinein die Beurteilung der Einhaltung von entsprechenden Sorgfaltsmaßstäben erfolgen kann (retrospektiver Ansatz). Deshalb muss gewährleistet sein, dass die KI-Systeme nicht nur vergangenheitsbezogen lernen, sondern offen für Regeländerungen (wie den Erlass neuer Gesetze oder die Änderung relevanter Rechtsprechung) sind. Insoweit können die rechtlichen Regelungen des „Privacy by Design“-Ansatzes nach Art. 25 DSGVO als für KI-Systeme zu verallgemeinernde Vorlage genutzt werden, soweit diese in kritischen Situationen eingesetzt werden (z.B. in der Medizin, aber ggf. auch bei der Steuerung wichtiger infrastruktureller Maßnahmen). Ob dann bei einer Nichtbeachtung eine Sanktionierung (wie in Art. 84 DSGVO für eine Nichtbeachtung des Art. 25 DSGVO vorgesehen) erfolgen sollte, wäre entsprechend der Sensibilität des Einsatzbereiches kritisch zu prüfen. Dies betrifft den gesamten Lebenszyklus¹⁹⁵ eines KI-Systems. Damit wird nur die an sich triviale Selbstverständlichkeit adressiert, dass das geltende Recht zu befolgen ist und für den Fall der Nichtbefolgung eine entsprechende Aufklärung möglich ist. Das ist auch deshalb notwendig, um die demokratische Basis unseres Rechtssystems auch für die Zukunft zu sichern und eine „Delegation“ der Setzung der im Alltag wirksam werdenden Normen auf die Programmierer und -betreiber zu verhindern.

Das Nutzungspotential der „künstlichen Intelligenz“ hat einen unmittelbaren Einfluss auf die Sorgfaltsmaßstäbe und damit auf die Grenze zwischen rechtmäßigen und rechtswidrigen Verhaltensweisen. Wenn beispielsweise eine Weiterentwicklung des oben erwähnten deep-learning-Systems zur Erkennung von Hautkrebs¹⁹⁶ in der Lage ist, bessere „Entscheidungen“ auf der Grundlage einer schnelleren und präziseren Interpretation von Ergebnissen zu erzielen, als dies menschliche Ärzte könnten, ist eine Bestimmung des erlaubten Risikos notwendig. Im Sinne einer Beobachtungspflicht sollte der Gesetzgeber regelmäßig evaluieren, inwieweit ein gesetzgeberisches Eingreifen erforderlich ist.

Eine Sanktionierung von KI-Systemen direkt ist derzeit nicht sinnvoll. Hierfür wäre zunächst die Etablierung entsprechender Mechanismen zur Verbreitung von rechtskräftig gegen KI-Systeme verhängten Sanktionen notwendig, die bei der Abarbeitung des Algorithmus (automatisch) mitberücksichtigt werden. Außerdem müsste unter dem Gesichtspunkt

der Generalprävention insbesondere bei dem durch eine Sanktion betroffenen KI-System eine Funktionalität implementiert sein, die dieses davon abhält, das sanktionierte Verhalten noch einmal zu begehen. Insoweit sollten für die Frage der Sanktionierung Parallelen nicht zum Menschen, sondern zu künstlichen Akteuren (wie den juristischen Personen) gezogen werden.

Die sich abzeichnende Einführung eines Unternehmensstrafrechts sollte dazu genutzt werden, das Verbund- und Vernetzungsrisiko beim Betrieb von autonomen Systemen durch Unternehmen zu berücksichtigen. Dabei könnte das kommende Unternehmenssanktionenrecht vorsehen, dass der Normadressat für entsprechende Verbandssanktionen funktional danach bestimmt wird, wer faktisch die Regeln für den Einsatz von KI-Systemen festlegt. Außerdem sollte sichergestellt werden, dass das Fehlverhalten autonomer Systeme direkt dem Verband zugerechnet werden kann. Beispielsweise könnten „KI-Sicherheitsbeauftragte“ als Leitungspersonen gefordert werden. Diese hätten dann zu beobachten und bei erkannten Abweichungen sicherzustellen, dass sich die eingesetzten KI-Systeme innerhalb der rechtlichen Grenzen bewegen. Hierfür ist zwingend eine effektive menschliche Aufsicht über die eingesetzten Systeme sicherzustellen. Dabei ist jeweils nach den Einsatzgebieten zu differenzieren, ob ein „Human on the Loop“, „Human in the Loop“ oder ein „Human in Command“ erforderlich ist.

¹⁹⁴ Das Fehlen einer dem Paul-Ehrlich-Institut vergleichbaren Behörde kritisiert Yuan, RW 2018, 477 (495).

¹⁹⁵ Vgl. oben zu Fn. 10.

¹⁹⁶ Vgl. oben zu Fn. 140.