weizenbaum
institut

Weizenbaum Series #2

Aufsatz

Risiken digitaler Systeme:

Robotik, Lernfähigkeit und Vernetzung als aktuelle Herausforderungen für das Recht

Herbert Zech

Januar 2020

Weizenbaum Series

Edited by

Weizenbaum Institute for the Networked Society –
The German Internet Institute

Project Coordination:
Wissenschaftszentrum Berlin für Sozialforschung
Reichpietschufer 50
10785 Berlin

Visiting Address:
Hardenbergstraße 32
10623 Berlin

Email: info@weizenbaum-institut.de
Web: www.weizenbaum-institut.de

Die Langzeitarchivierung dieser Reihe wird durch das Social Science Open Access Repository und den DOI-Registrierungsservice für Sozial- und Wirtschaftsdaten in Deutschland da|ra sichergestellt.

DOI 10.34669/wi.ws/2

Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen: 16DII111, 16DII112, 16DII113, 16DII114, 16DII115, 16DII116, 16DII117 – „Deutsches Internet-Institut“).

Weizenbaum Series #2
Aufsatz

Risiken digitaler Systeme: Robotik, Lernfähigkeit und Vernetzung als aktuelle Herausforderungen für das Recht

Herbert Zech

Januar 2020

Inhaltsverzeichnis

A. Einleitung

Digitale Systeme stehen im Mittelpunkt aktueller rechtlicher und politischer Diskussionen. Stichworte wie Künstliche Intelligenz oder Autonomie werden dabei nicht immer einheitlich verwendet. Dieser Beitrag möchte die wichtigsten Begriffe klären und einen Überblick über die entscheidenden aktuellen Entwicklungen im Bereich der Informationstechnologie geben.[1] Er ist aus dem Blickwinkel eines Juristen geschrieben, beschäftigt sich aber nur mit den technischen Aspekten.

B. Digitale Systeme leisten technische Informationsverarbeitung

Spricht man von digitalen Systemen, so geht es um den Bereich der Informationstechnologie (IT). Ihr zentrales Merkmal ist die Informationsverarbeitung durch technische Systeme. Digitale Systeme sind künstliche bzw. technische informationsverarbeitende Systeme (I.). Während der Begriff Informationstechnologie nach wie vor trennscharf verwendet werden kann, ist dies für den Begriff der Künstlichen Intelligenz schwieriger. Wichtiges Merkmal beider ist jedenfalls die Automatisierung.

Informationsverarbeitende Systeme zeichnen sich dadurch aus, dass sie Information eingegeben bekommen oder selbsttätig aufnehmen (Informationseingabe), verarbeiten und ausgeben (II.). Je nach Art der Eingabe und der Ausgabe können Systeme mit verschiedenen Schwerpunkten unterschieden werden, insbesondere Datenanalyse- und Steuerungssysteme.

Informationsverarbeitende Systeme lassen sich zudem danach unterscheiden, ob die Information symbolisch (zeichenhaft) oder implizit repräsentiert wird (III.). Während digitale Systeme grundsätzlich Information binär codierter Form (Bits und Bytes) verarbeiten, kann auf höherer Ebene zwischen der symbolischen Repräsentation in Form von Programmen, Daten etc. und der impliziten Repräsentation in Form von Netzwerken (künstliche neuronale Netze) unterschieden werden.

Die Abgrenzung digitaler Systeme kann sowohl auf virtueller als auch auf physischer Ebene erfolgen (IV.). Während die Robotik dafür sorgt, dass digitale Systeme unmit-

1 Der Beitrag ist eine ausführlichere Version des technischen Einführungskapitels für das Gutachten A zum 73. Deutschen Juristentag „Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?“.

telbare physische Auswirkungen haben können, bewirkt die zunehmende Vernetzung umgekehrt, dass eine Abgrenzung auf der physischen Ebene immer schwieriger wird. Digitale Systeme entwickeln sich nach wie vor rasant. Ihre zunehmende Komplexität und Leistungsfähigkeit folgen einem seit langem bestehenden Trend (Moore's Law). Komplexität ist ein Kennzeichen moderner Technik und als solche nichts Besonderes. Die Leistungsfähigkeit digitaler Systeme hat aber in den letzten Jahren einen Wendepunkt erreicht, der eine Vielzahl neuer Anwendungen möglich macht und dafür sorgt, dass digitale Systeme sämtliche Lebensbereiche durchdringen (V.).

I. Informationsverarbeitung durch natürliche und technische Systeme

1. Informationsverarbeitung als zentrales Merkmal

Das zentrale Merkmal aller digitalen Systeme (IT-Systeme) stellt die Informationsverarbeitung dar. Darin liegt auch die Gemeinsamkeit mit natürlichen informationsverarbeitenden Systemen (Intelligenzen). Diese Gemeinsamkeit hat zur Entwicklung der Kognitionswissenschaften (cognitive sciences) beigetragen, die natürliche und technische Informationsverarbeitung gleichermaßen untersuchen. In einem Standardlehrbuch heißt es dazu: „The most fundamental driving assumption of cognitive science is that minds are information processors."[2]

Der Begriff der Information selbst soll hier nicht weiter vertieft werden.[3] In ihrer grundlegendsten Bedeutung lässt sich Information als „beseitigte Ungewissheit"[4] verstehen, was allerdings bereits einen Verstand bzw. ein Bewusstsein voraussetzt. Indem Information durch Zeichen repräsentiert wird, kann sie nicht nur kommuniziert, sondern auch maschinell verarbeitet werden. Zur klassischen zeichenhaften Repräsentation ist mit dem Aufkommen leistungsfähiger künstlicher neuronaler Netze die implizite Repräsentation nach dem Vorbild natürlicher neuronaler Netze hinzugekommen (dazu unter III.).

---

2 *Bermudez*, Cognitive Science, 2. Aufl. 2014, S. 2.

3 Dazu *Zech*, Information als Schutzgegenstand, 2012, S. 13 ff.

4 *Klimant/Piotraschke/Schönfeld*, Informations- und Kodierungstheorie, 3. Aufl. 2006, S. 12; *Illik*, Formale Methoden der Informatik, 2009, S. 101.

2. Informationstechnologie (IT) und Automatisierung

Der für die künstliche bzw. maschinelle Verarbeitung von Information bereits sehr früh eingeführte Begriff der Informationstechnologie erweist sich nach wie vor als gut geeignet, um diesen Technikbereich zu kennzeichnen.[5] Der Begriff des digitalen Systems bzw. der Digitalisierung stellt die Verbindung zur ursprünglich menschlichen Tätigkeit des Rechnens her. Ebenso werden Begriffe wie Computer (computing) oder Rechner verwendet. Digital bedeutet ursprünglich „die Finger benutzend" und verweist auf das Zählen und Rechnen, das mit zeichenhaft repräsentierter Information möglich wird.

Dazu erforderlich ist eine Automatisierung in der Grundbedeutung der Selbsttätigkeit (ohne menschliche Mitwirkung). Maschinelle Verarbeitung von zeichenhaft repräsentierter Information bedeutet automatisiertes logisches Schließen. Auch Künstliche Intelligenz (dazu unter 3.) wird daher als „stetiges Voranschreiten der Automatisierung"[6] beschrieben.

Automatisierung ist damit der Oberbegriff für den selbsttätigen Ablauf von Prozessen und bezieht sich nicht nur auf die Steuerung von Hardware, sondern der Informationsverarbeitung insgesamt. Zu den Merkmalen, die jede maschinelle Informationsverarbeitung ausmachen, gehören Steuerung, Speicher, Rechner und die Verwendung elektronischer Schaltkreise.[7] Die Umsetzung in immer leistungsfähigeren elektronischen Schaltkreisen macht die vielfältigen aktuellen Anwendungen (dazu unter V.) überhaupt erst möglich. Mechanische Umsetzungen sind zu langsam, Umsetzungen in anderen physikalischen Phänomenen, insbesondere optische Rechner, stehen erst am Anfang.

Mit der Automatisierung verknüpft ist auch der Begriff des Programms. Es sorgt für einen Ablauf nach vorgegebenen Regeln (auch, wenn diese unter Umständen sehr komplex sind). Unter C. wird zu zeigen sein, dass Informationsverarbeitung bei lernfähigen Maschinen auch ohne festes Programm möglich ist.

---

5 Zum Technikbegriff *Zech*, in FS Bodewig, 2018, S. 137, 150 ff.

6 *Kaplan*, Künstliche Intelligenz, 2015, S. 32.

7 *Ceruzzi*, Computing: A Concise History, 2012, S. 4 ff. Ebd., S. 9: „Control, storage, calculation, the use of electrical or electronic circuits: these attributes, when combined, make a computer."

3. Begriff der Künstlichen Intelligenz

Der Begriff der Künstlichen Intelligenz resultiert aus dem Vergleich natürlicher und künstlicher (technischer) Informationsverarbeitung. Bereits 1955 wurde in einem Forschungsantrag für die sogenannte Dartmouth Conference von einer Vermutung gesprochen, dass jeder Aspekt des Lernens oder jedes andere Merkmal von Intelligenz grundsätzlich so genau beschrieben werden könne, dass eine Maschine in der Lage ist, diese zu simulieren.[8] Die klassische Definition Künstlicher Intelligenz besteht also in der Nachahmung natürlicher Intelligenz.

Gleichwohl ist der Begriff der Künstlichen Intelligenz mit großen Ungenauigkeiten befrachtet, weshalb er eher als Schlagwort verwendet wird.[9] Die ihm seit seiner Entstehung zugrundeliegende Idee ist, dass natürliche Intelligenz durch Technik nachgeahmt werden kann. Während dies zunächst durch symbolisch repräsentierte Informationsverarbeitung erreicht werden sollte, beruhen die spektakulären aktuellen Erfolge zumindest teilweise auf implizit in Netzwerken repräsentierter Informationsverarbeitung (dazu unter III.). Diese ist der natürlichen Informationsverarbeitung tatsächlich ähnlicher als klassische Rechner.

Der Begriff der Künstlichen Intelligenz bleibt aber problematisch.[10] Er birgt die Gefahr einer unhinterfragten Gleichstellung digitaler Systeme mit natürlicher oder gar menschlicher Intelligenz.[11] Zur Kennzeichnung höherer bzw. komplexerer Aufgaben der Informationsverarbeitung eignet er sich aber durchaus. Entsprechend sind heute viele Teilbereiche der Forschung in der Informationstechnologie durch Aufgaben gekennzeichnet, die Menschen oder Tiere lösen können und digitale Systeme lösen sollen (siehe unter V.).

---

8 *McCarthy u.a.*, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955 (verfügbar unter http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, zuletzt aufgerufen am 10.10.2019): „We propose that a […] study of artificial intelligence be carried out […]. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

9 Siehe aber auf europäischer Ebene die ausführliche Definition der Hochrangigen Expertengruppe für künstliche Intelligenz, Eine Definition der KI: Wichtigste Fähigkeiten und Wissenschaftsgebiete, 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60664 (zuletzt aufgerufen am 10.10.2019).

10 *Herberger*, NJW 2018, 2825 ff.

11 *Balkin*, 78 Ohio St. L.J. 1217 (2017), 1223: „homunculus fallacy".

II. Verschiedene Formen der Ein- und Ausgabe von Information

Ihrer Grundstruktur nach verfügen informationsverarbeitende Systeme immer über die Möglichkeit der Informationsaufnahme bzw. -eingabe, der eigentlichen Informationsverarbeitung und der Informationsausgabe (input, processing, output). Im Begriff des Systems ist impliziert, dass es von der Umwelt zumindest gedanklich abgegrenzt werden kann, auch wenn die zunehmende Vernetzung, wie noch zu zeigen sein wird, gerade ein Merkmal fortgeschrittener informationsverarbeitender Systeme ist.

1. Grundmodell: Eingabe, Verarbeitung, Ausgabe

Das Grundmodell „input – processing – output" findet sich in allen Systemen der Informationstechnologie. Sowohl die Eingabe als auch die Ausgabe können virtuell (Ein- und Ausgabe von Zeichen) oder physisch (Aufnahme durch Sensoren, Ausgabe durch Aktuatoren) erfolgen. Die besondere Bedeutung physischer Systeme wird noch gesondert dargestellt, sowohl von Robotern, die sich durch Ausgabe in Form von Hardwaresteuerung auszeichnen und häufig mit Sensoren ausgestattet sind (dazu unter B.), als auch von Alltagsgegenständen, die mit Sensoren ausgestattet sind und so zum Bestandteil von cyber-physikalischen-Systemen werden (dazu unter C.).[12]

Ein besonders generalisiertes System (der Begriff der „starken" künstlichen Intelligenz ist eher verwirrend) wäre sowohl zu selbständiger Eingabe durch Wahrnehmung (Sensorik) als auch zur Ausgabe durch „Handlung" (Aktuatoren) fähig. Je nach Generalisierungs- bzw. Spezialisierungsgrad in wenigen oder zahlreichen Situationen lassen sich Systeme mit mehr oder weniger Freiheitsgraden unterscheiden. Durch den Einsatz lernfähiger Systeme (dazu unter C.) kann maschinelle Informationsverarbeitung heute einen wesentlich höheren Generalisierungsgrad erreichen als noch vor einem Jahrzehnt.

Die Informationseingabe und -ausgabe kann unterschiedliche Formen annehmen. Klassischerweise stehen an beiden Enden Menschen, was eine Ein- bzw. Ausgabe in Form von für Menschen wahrnehmbaren Zeichen voraussetzt. Zunehmend können Systeme jedoch selbständig Information aufnehmen (Erfassung von Rohdaten, Big Data). Auch können sie, wie bereits erwähnt, statt Information an Menschen auszu-

12 Vgl. *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 227 f.: „Mund-Kopf-Körper-Architektur".

geben, auch steuernd in physikalische Abläufe eingreifen. Die für die jeweilige Informationsverarbeitung bzw. das eingesetzte System verwendeten Begriffe unterscheiden sich jeweils danach, wo der Schwerpunkt liegt. Als wichtige Pole sollen die Datenanalyse und die Steuerungstechnik erwähnt werden.

2. Datenanalyse

Liegt der Fokus beim Einsatz digitaler Systeme auf der Analyse, entspricht dies dem Bereich der Data Sciences (Big Data). Hauptaufgabe ist es dabei, Muster in Daten zu erkennen. Data-Science-Studien lassen sich allgemein als Prozess aus vier Schritten darstellen: Aufbereitung der Daten, Auswahl der geeigneten Algorithmen, Optimierung der Parameter der Algorithmen, Ableitung von Modellen und deren Evaluierung und Validierung mit Auswahl des besten Modells.[13]

Kennzeichen von Datenanalysen ist, dass die Ausgabe der Informationsverarbeitung grundsätzlich als Information an Menschen erfolgt (eine automatisierte Verwendung ist aber möglich). Gelegentlich spricht man von „veredelten" Daten, allgemein kann man aber auch vom Ergebnis einer Datenanalyse sprechen. Die Information kann schließlich auch darin bestehen, dass eine automatisierte Entscheidung getroffen wird.

Big Data zeichnet sich dadurch aus, dass die Eingabe nicht unmittelbar durch Menschen erfolgt, sondern durch Sensoren, die zunehmend in der physischen Welt verbreitet sind und damit die Erzeugung großer Datenmengen (Rohdaten, Inputdaten) ermöglichen.[14]

Eine neuartige Situation würde entstehen, wenn der Input ohne Eingabe durch den Menschen direkt aus dem menschlichen Gehirn erfolgen könnte. Die entsprechende Technologie der Brain Computer Interfaces befindet sich aber noch im Forschungs- und Entwicklungsstadium.

13 *Ng/Soo*, Data Science – Was ist das eigentlich?, 2018, S. 1 f.

14 Dazu grundlegend *Mayer-Sch*önberger/*Cukier*, Big Data – A Revolution That Will Transform How We Live, Work, and Think, 2013.

3. Steuerung

Liegt der Schwerpunkt auf der Steuerung, so erfolgt die Ausgabe in Form von Steuerungsbefehlen. Steuerung bedeutet zwar nicht notwendig die Steuerung von Hardware, diese stellt aber einen wichtigen Bereich dar. So werden selbstfahrende Fahrzeuge aktuell mit großem Aufwand entwickelt und auch andere Bereiche der Robotik könnten in mehr oder weniger naher Zukunft im Alltag Verbreitung finden (dazu unter B.). Roboter verfügen in der Regel sowohl über eigene Sensoren als auch Aktuatoren, daher auch die Definition als Maschinen, die wahrnehmen, denken und handeln (dazu unter B.I). Die Sensoren in Verbindung mit Informationsverarbeitung und Steuerung von Aktuatoren ermöglichen die Automatisierung von Hardware. Automatisierung bedeutet hier, dass Hardware ohne menschlichen Eingriff „funktioniert" bzw. auf Reize von außen (außerhalb des Systems) reagiert.

III. Symbolische oder implizite Repräsentation von Information

Eine wichtige Unterscheidung, die auch für die Lernfähigkeit von Maschinen von Bedeutung ist, betrifft die Art und Weise, wie Information in dem informationsverarbeitenden System repräsentiert oder – mit anderen Worten – „enthalten" ist: Information kann symbolisch repräsentiert sein oder implizit.[15]

1. Symbolische Repräsentation: logisches Schließen, klassische Programmierung, Algorithmen

Klassische Programmierung beruht auf symbolischer Repräsentation (wie auch menschliche Sprache oder Schrift), d.h. es gibt eine klare Zuordnung von Zeichen zu Bedeutung bzw. technischer Funktion. Die Informationsverarbeitung erfolgt dann durch logische Operationen mit den Zeichen. Anders dagegen etwa natürliche Gehirne, aber auch künstliche neuronale Netze (dazu unter 2.): Hier gibt es keine klar definierte Zeichenebene (nur Struktur und Bedeutung), jedenfalls bei der eigentlichen Verarbeitung (durchaus bei der Ein- und Ausgabe).

15 *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 224 f.

Die klassische Künstliche Intelligenz (Good Old-Fashioned AI, GOFAI) beruht auf einer vollständigen symbolischen Repräsentation der zu verarbeitenden Information. Der bereits erwähnte Förderantrag von *Carthy u.a.* macht dies deutlich: „every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it".[16] Die Beschreibung durch den Programmierer erzeugt symbolisch repräsentierte Information.

Die symbolische Repräsentation von Information ermöglicht auch die Abgrenzung von Information auf einer eigenen Zeichenebene, die zwischen Bedeutungsebene und Strukturebene (Struktur des informationsverarbeitenden Systems) tritt. Dadurch kann Information wiederum von einem einzelnen informationsverarbeitenden System abgelöst und als eigenständiges Objekt aufgefasst werden.[17] Daten sind symbolisch repräsentierte maschinenlesbare Information. Die Bedeutung kann auch in der maschinensteuernden bzw. computersteuernden Funktion liegen (funktionale Information).

Der entscheidende Effekt für die Informationsverarbeitung besteht darin, dass Zeichen automatisiert verarbeitet werden können: symbolische Repräsentation erlaubt maschinelles logisches Schließen. Dies bedeutet, dass symbolisch repräsentierte Information durch einen Algorithmus verarbeitet werden kann.

Ein *Algorithmus* ist eine ihrerseits symbolisch repräsentierbare Handlungsanweisung. Übliche Definitionen[18] lauten etwa „a recipe that, if followed, guarantee[s] a solution"[19] oder „Abfolgen von klar vorgegebenen, eindeutig definierten Handlungsschritten"[20]. Die klassische künstliche Intelligenz ging davon aus, dass sich intelligentes Verhalten durch Algorithmen beschreiben lässt und alle durch Intelligenz lösbare Aufgaben auch mit Hilfe von Algorithmen lösen lassen, sie hatte die „algorithmische Abbildung intelligenter Verhaltensmuster"[21] als Grundlage. Dies führte jedoch zwingend zu Problemen wegen der hohen Komplexität der Aufgaben (insbesondere wegen der Komplexität der Umwelt), die wiederum eine nicht beherrschbare Informationsmenge bei den Algorithmen erfordert hätte. Die Lösung dafür lag in der Entwicklung lernfähiger informationsverarbeitender Systeme, die unter C. ausführlich dargestellt werden.

---

16 *McCarthy u.a.*, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955 (verfügbar unter http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, zuletzt aufgerufen am 10.10.2019).

17 *Zech*, Information als Schutzgegenstand, 2012, S. 38 ff.

18 Dazu *Reichwald/Pfisterer*, CR 2016, 208, 209; *Wischmeyer*, AöR 143 (2018), 1, 4; *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, S. 17; *Zech*, ZfPW 2019, 198, 199; *ders*., Information als Schutzgegenstand, 2012, S. 21.

19 *Ceruzzi*, Computing: A Concise History, 2012, S. 33.

20 *Reichwald/Pfisterer*, CR 2016, 208, 209.

21 *Reichwald/Pfisterer*, CR 2016, 208, 211.

2. Implizite Repräsentation: neuronale Netze

Wie bereits angesprochen wird Information in natürlichen Systemen grundsätzlich implizit repräsentiert. Insbesondere in neuronalen Netzwerken ist die Information im Gesamtzustand des Netzwerks enthalten und nicht in Form von Zeichen. Dieses Prinzip wird in künstlichen neuronalen Netzen nachgeahmt. Die Regeln, nach denen die Informationsverarbeitung in künstlichen neuronalen Netzen erfolgt, lassen sich nur als Trainingszustand in Form von Gewichtungen (weights) der einzelnen Knoten in einem bestimmten Netz beschreiben (dazu unter C.VI).

Das Verhalten eines Netzes insgesamt (also die Regeln, nach denen die Informationsverarbeitung erfolgt) ist kein Algorithmus, wohl aber kommen Algorithmen als Elemente der Architektur von künstlichen Netzen zum Einsatz. Zudem besteht eine der aktuellen Forschungsaufgaben darin, die symbolische Beschreibung des Verhaltens von Netzen (mit mehr oder weniger hoher Zuverlässigkeit) zu ermöglichen (Explainable AI, XAI).

Zudem sind Übergänge zwischen impliziter und symbolischer Repräsentation möglich, etwa wenn ein natürliches Gehirn sprachlich kommuniziert. Auch bei künstlichen neuronalen Netzen können die Eingabe oder die Ausgabe durch symbolisch repräsentierte Information erfolgen. Dies ermöglicht die Konstruktion hybrider Systeme, die für die Praxis von besonderer Bedeutung sind.

3. Unterschiedliche Anwendungsbereiche

Die Art der Repräsentation von Information hat Auswirkung auf die durch die Informationsverarbeitung lösbaren Aufgaben. Dadurch ergeben sich unterschiedliche Anwendungsbereiche. *Kaplan* formuliert dies so: „Generell ist die symbolische Argumentation besser für Probleme geeignet, bei denen abstraktes Schlussfolgern gefragt ist – das Machine Learning eignet sich dagegen für Situationen, in denen es auf sensorische Wahrnehmung oder das Extrahieren von Mustern aus stark verrauschten Daten ankommt.“[22]

Interessant ist aber gerade der Wechsel zwischen den beiden Typen, insbesondere die symbolische Ausgabe bei künstlichen neuronalen Netzen und die Forschung an erklärbarer künstlicher Intelligenz. Sie spielen auch für das Haftungsrecht eine große Rolle,

22 *Kaplan*, Künstliche Intelligenz, S. 54.

da Erklärbarkeit bzw. Beschreibbarkeit des Verhaltens (auch wenn sie nur mit einer gewissen statistischen Wahrscheinlichkeit gewährleistet werden kann) die Aufklärung von Kausalketten und damit ex ante Vorhersehbarkeit, bei Einflussmöglichkeiten Beherrschbarkeit und ex post Aufklärbarkeit ermöglichen kann (dazu unter D.).

IV. Virtuelle und physische Systeme: digitale Systeme und ihre Implementierung

1. Systembegriff: Abgrenzbarkeit

Im Begriff des Systems ist die Abgrenzbarkeit gegenüber der Umwelt enthalten.[23] Gerade bei digitalen Systemen fällt jedoch die Abgrenzung nicht immer leicht, vielmehr ist die zunehmende Vernetzung und damit erschwerte Abgrenzung ein wichtiger Aspekt der aktuellen technischen Entwicklung. Teilweise wird schon von einer Entwicklung zu einer Umgebungsrobotik bzw. „Infrastrukturrobotik“[24] gesprochen, d.h. die gesamte Lebensumwelt wird zu einem digitalen System (dazu unter D.)
Nach wie vor ist aber eine Abgrenzung möglich und muss es auch sein, um so noch Verantwortlichkeit zu bestimmen. Nur ein definierbares System kann einen Entwickler, Betreiber oder Benutzer haben. Die Unterschiede in der Beherrschung des Systems von der umfassenden Herstellung und Steuerung bis hin zum bloßen An- und Ausschalten oder dem Fehlen jeder Einflussmöglichkeit werden noch genauer dargestellt (unter E.III).

2. Hardwareebene und virtuelle Ebene

Die Besonderheit digitaler Systeme liegt darin, dass sie grundsätzlich sowohl auf Hardwareebene (als physikalische Struktur) als auch auf virtueller Ebene (als zeichenhafte Repräsentation der Regeln, nach denen die Informationsverarbeitung abläuft) abgegrenzt werden können. Darin liegt auch die besondere Herausforderung für Haftungsregeln, die an den Produktpfad anknüpfen, der herkömmlich auf körperliche Risikoquellen ausgelegt ist (dazu unter D.).

23 System wird hier einfach als Ausschnitt der Wirklichkeit verstanden, d.h. es kommt vor allem auf die Abgrenzbarkeit an, vgl. *Zech*, Information als Schutzgegenstand, 2012, S. 15.

24 *Klaus Mainzer*, Vortrag am 5.10.2015, Universität Passau. Vgl. *ders*., Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 155 ff.

Digitale Systeme verfügen immer über einen Hardwareaspekt, d.h. die Implementierung als physikalisches System bzw. Ausschnitt der naturwissenschaftlich beschreibbaren Wirklichkeit, und einen Softwareaspekt bzw. informationellen Aspekt, der sich losgelöst von der physikalischen Implementierung durch Zeichen beschreiben lässt. Diese Loslösung lässt sich als Virtualisierung bezeichnen. Technisch wird sie dadurch verstärkt, dass vielfach auf der Softwareebene abgrenzbare Systeme keinem klar abgrenzbaren Hardwaresystem mehr zugeordnet sind (was vielfach auch als Cloud bezeichnet wird). Obwohl immer irgendwo eine physikalische Implementierung vorliegen muss, kann diese auch verteilt sein und ist häufig praktisch nicht mehr bestimmbar. Damit bleibt die informationelle Ebene als wichtiger Abgrenzungs- und Beherrschungsfaktor. Hier besteht umgekehrt das Problem, dass durch das Aufkommen lernfähiger Systeme auch die Beschreibung der informationellen Seite erschwert wird. Algorithmen können als Abfolge klar definierter Handlungsanweisungen zeichenhaft beschrieben werden. Anders ist es, wenn die Regeln, nach denen die Informationsverarbeitung abläuft, nur noch als trainiertes Netz zur Verfügung stehen (Beschreibung in Form von Gewichtungen der einzelnen Knoten). Hier fehlt eine klare Zuordnung von Zeichen und Bedeutung wie sie für klassische Algorithmen bzw. Software, die auf symbolischer Logik aufbaut, kennzeichnend ist. Digitale Systeme werden daher nicht nur virtueller, sondern auch schwerer beschreibbar.
Schließlich führt der dritte Aspekt, die zunehmende Verbreitung von Robotik, also Software-Hardware-Kopplung, dazu, dass digitale Systeme trotz Virtualisierung mehr und nicht weniger Einfluss auf die physikalische Welt nehmen können. Dies ist wiederum für das Haftungsrecht von erheblicher Bedeutung, da jeder unmittelbare Einfluss auf die physikalische Welt auch das Potential einer Schädigung körperlicher Gegenstände mit sich bringt (dazu unter B.).

3. Auswirkung auf verschiedenen Stufen der Wertschöpfungskette

Die Physikalität oder Virtualität digitaler Systeme kann sich je nach Stufe der Wertschöpfungskette ändern. Ein Wechsel zwischen den Ebenen kann mehrfach erfolgen. So könnte etwa zunächst eine lernfähige Software vertrieben werden, die in ein Hardwaresystem eingebaut wird. Dieses wird trainiert. Der fertige Trainingszustand kann theoretisch wieder unkörperlich vermarktet und von anderen Herstellern in physikalische Systeme eingebaut werden. Selbst nach dem Erwerb solcher Systeme durch die

Endnutzer bleibt ein unkörperlicher Austausch der Regeln, nach denen die Informationsverarbeitung abläuft, möglich, etwa in Form eines Updates oder eines Upgrades. Auf den Stufen der Entwicklung, Herstellung, Vermarktung, Anwendung/Betrieb oder Nutzung kommen damit unabhängig voneinander grundsätzlich sowohl körperliche als auch unkörperliche Verursachungspfade in Betracht.

V. Einsatzbereiche digitaler Systeme

Mit der technischen Entwicklung nehmen die Aufgaben, die digitale Systeme bewältigen können, und damit auch die Einsatzbereiche stetig zu. Die Informationstechnologie mit ihren aktuellen Entwicklungen stellt eine transformative Technologie dar.

1. Typische Aufgaben, die digitale Systeme lösen können

Wie bereits erwähnt ergeben sich mögliche Aufgaben, die digitale Systeme lösen können bzw. sollen, insbesondere aus dem Vergleich mit natürlicher Intelligenz. Der Umfang an Aufgaben, an deren Bewältigung durch digitale Systeme geforscht wird, ist entsprechend breit. *Flasiński* zählt folgende "application areas of AI systems" auf: perception and pattern recognition (Wahrnehmung und Mustererkennung), Knowledge Representation (Wissensrepräsentation), Problem Solving (Problemlösung), Reasoning (Schlussfolgern), Decision Making (Entscheiden), Planning (Planung), Natural Language Processing (Verarbeitung natürlicher Sprache), Learning (Lernen), Manipulation and Locomotion (Robotik), Social Intelligence (soziale Intelligenz), Emotional Intelligence and Creativity (emotionale Intelligenz und Kreativität).[25]
Zur Veranschaulichung der Schwerpunkte in der aktuellen Forschung lässt sich auch auf die Themen der 33. AAAI Conference on Artificial Intelligence (AAAI 2019) verweisen: Unter anderem finden sich dort Computersehen, Constraint-Probleme, Handlungsplanung, heuristische Suche und Optimierung, kognitive Modellierung, maschinelles Lernen, Multiagentensysteme, Robotik, Spieltheorie, Verarbeitung natürlicher Sprache, Wissensrepräsentation und Schlussfolgerung.[26]

---

25 *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 223 ff.

26 Für den Hinweis danke ich *Malte Helmert*. Vgl. https://aaai.org/Conferences/AAAI-19/ (zuletzt aufgerufen am 10.10.2019).

Als zentrale Aufgaben, für die digitale Systeme bereits eingesetzt werden, nennt *Stiemerling* Mustererkennung, Maschinelles Lernen, Expertensysteme (was der Aufgabe der Wissensrepräsentation entspricht) und Maschinelles Planen und Handeln.[27]

2. Einsatzbereiche, Digitalisierung als „horizontales" Phänomen

Einsatzbereiche für digitale Systeme finden sich mittlerweile in nahezu allen Wirtschaftsbereichen.[28] Zu den wichtigsten Bereichen, die eine Vorreiterrolle einnehmen, gehören automatisierte Fahrzeuge[29] und der Gesundheitsbereich.[30]

Die eingesetzten Systeme reichen von klassischen Schachprogrammen bis zu AlphaGo (von DeepMind),[31] von klassischen Expertensystemen für Ärzte bis zur Krebs-Diagnostik mit Hilfe maschinellen Lernens,[32] von einfachen Fahrassistenzsystemen bis zum vollautomatisierten Fahrzeug (das noch nicht existiert).

Die Entwicklung selbstfahrender Fahrzeuge ist für den Automobilsektor von enormer Bedeutung. Aktuell werden bereits Fahrzeuge mit einem gewissen Grad an Automatisierung eingesetzt, jedoch noch keine vollautomatisierten Fahrzeuge. § 1a StVG trägt dem bereits Rechnung, indem er Kraftfahrzeuge mit hoch- oder vollautomatisierter Fahrfunktion reguliert. Es existieren auch nationale und internationale technische Standards, die verschiedene Grade an Automatisierung definieren. Selbstfahrende Fahrzeuge entsprächen level 5 („full automation" nach SAE/Society of Automotive Engineers bzw. „autonom" nach BASt/Bundesanstalt für Straßenwesen).[33] Der Begriff vollautomatisiert ist hier passender als autonom, da Autonomie bei digitalen Systemen häufig auch mit Lernfähigkeit gleichgesetzt wird (dazu unter C.).

---

27 *Stiemerling*, CR 2015, 762 ff.

28 Beispiele etwa bei *Borges*, NJW 2018, 977, 979; *Eidenmüller*, The Rise of the Robots and the Law of the Humans, SSRN Nr. 2941001, S. 2 f.; *Lohmann*, AJP 2017, 152, 155 ff. (zu Robotik).

29 Maurer u.a. (Hrsg.), Autonomes Fahren, 2015.

30 *Schönberger*, IJLIT 2019, 171 ff.

31 https://deepmind.com/ (zuletzt aufgerufen am 10.10.2019).

32 Dazu etwa Interview mit *Klaus-Robert Müller*, https://ki-berlin.de/blog/article/prof-dr-klaus-robert-mueller-tu-berlin/ (zuletzt aufgerufen am 10.10.2019).

33 Dazu *Maurer*, in ders. u.a. (Hrsg.), Autonomes Fahren, 2015, S. 1, 2 ff.; *Hey*, Die außervertragliche Haftung des Herstellers autonomer Fahrzeuge bei Unfällen im Straßenverkehr, 2019, S. 8 ff.; *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 169; *ders.*, ZfPW 2019, 198, 199 f. Zu Fahrassistenzsystemen bereits *Bewersdorf*, Zulassung und Haftung bei Fahrassistenzsystemen im Straßenverkehr, 2005, S. 28 ff.

Rein virtuelle Systeme, die erfolgreich komplexe Aufgaben leisten, werden unter anderem zur Spracherkennung und Übersetzung eingesetzt. Hier wurden durch die Verwendung lernfähiger Systeme große Qualitätssprünge erzielt.[34]
Im Bereich der Robotik finden sich mögliche Anwendungsfelder neben dem bereits genannten Verkehrssektor in der Pflege, im Haushalt und natürlich auch im militärischen Bereich.[35]

3. Digitale autonome Systeme als transformative Technologie

Digitale Systeme halten Einzug in nahezu alle Bereiche menschlichen Handelns. Man kann von einem „horizontalen" Phänomen sprechen, einer grundlegenden Technologie, die als Kulturtechnik der Sprache, Schrift oder Büchern vergleichbar ist.
Wegen ihrer Leistungsfähigkeit und der Durchdringung aller Lebensbereiche haben digitale Systeme mit ihren aktuellen Entwicklungen das Potential, tiefgreifende Veränderungen der jeweiligen menschlichen Verhaltensweisen herbeizuführen. Es handelt sich um eine transformative Technologie.[36] Aus juristischer Perspektive ist dies deshalb interessant, weil damit auch das Potential verbunden ist, gesellschaftliche Wertvorstellungen zu verändern, was wiederum Druck auf gesetzliche Regelungen ausüben kann. Rein deskriptiv handelt es sich jedenfalls um eine sehr wirkmächtige technische Entwicklung. Dies wird gut durch die These vom zweiten Maschinenzeitalter nach *Brynjolfsson/McAfee* illustriert:[37] Der Gewinn an Leistungsfähigkeit in der Informationsverarbeitung hat dieser These zufolge in den letzten Jahren einen Punkt erreicht, der

34 Dazu instruktiv *Lewis-Kraus*, The Great A.I. Awakening, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html (zuletzt aufgerufen am 10.10.2019).

35 Dazu *Calo/Froomkin/Kerr*, Robot Law, 2016, S. 333 ff.

36 Europäische Kommission, Mitteilung COM(2018) 237 final, 1: „As with any transformative technology, some AI applications may raise new ethical and legal questions, for example related to liability or potentially biased decision-making." In der deutschen Fassung: „Wie jede revolutionäre Technologie können einige KI-Anwendungen neue ethische und rechtliche Fragen aufwerfen, die etwa die Haftung oder potenziell parteiische Entscheidungen betreffen." Zum Begriff der transformativen Technologie Nuffield Council on Bioethics, Genome editing: an ethical review, 2016, 12, 26; *Fateh-Moghadam*, medstra 2017, 146 (148); *ders.*, ZStW 131 (2019). Vgl. *Calo*, 103 Cal. L. Rev. 513 (2015), 528 zur Robotik.

37 *Brynjolfsson/McAfee*, The Second Machine Age, 2014, S. 112 f.: „Weil es die exponentiellen, digitalen und neu kombinierenden Kräfte des zweiten Maschinenzeitalters den Menschen ermöglicht haben, zwei der wichtigsten einmaligen Ereignisse unserer Geschichte herbeizuführen: die Entstehung echter, nutzbarer künstlicher Intelligenz (Artificial Intelligence oder AI) und die Anbindung der meisten Menschen auf der Erde an ein gemeinsames digitales Netz." Dazu *Frese*, NJW 2015, 2090 ff.

demjenigen vergleichbar ist, an dem die Energieerzeugung (bzw. -umwandlung) mit der Verbesserung der Dampfmaschinen durch James Watt stand. Dieser Punkt wird zu einem Wendepunkt, an dem die kontinuierliche Verbesserung der Leistungsfähigkeit zu einem sprunghaften Anwachsen der Anwendungsmöglichkeiten führt (was bei den Dampfmaschinen zu einem wesentlichen Faktor der industriellen Revolution wurde).

C. Robotik: von der virtuellen in die reale Welt

Entwicklungen in der Robotik waren der erste Anstoß für ein Aufleben der haftungsrechtlichen Diskussion über IT-Systeme. Während die schon lange diskutierte Haftung für Software ein Spezialthema blieb,[38] führten Roboter dazu, dass unmittelbare physische Schädigungen in immer mehr Lebensbereichen denkbar wurden. Technische Grundlage sind Entwicklungen, die IT-gesteuerten Hardwaresystemen mehr Freiheitsgrade verschaffen (dazu unter II.) und so zu einer Verbreitung in allen räumlichen Lebensbereichen beitragen (III.). Während einfache ortsfeste Roboter am Anfang der Entwicklung standen, könnten humanoide Roboter und Infrastrukturrobotik die Zukunft prägen.

I. Roboter als unmittelbar durch digitale Systeme gesteuerte Hardware (Ausgabe erfolgt unmittelbar an Aktuatoren)

Roboter sind IT-gesteuerte Maschinen.[39] Entscheidend ist, dass sie als Hardwaresysteme definiert und eingesetzt werden können. Die Ausgabe der Informationsverarbeitung erfolgt unmittelbar als Steuerung von Hardware. Dementsprechend lassen sich

38 Zur Diskussion um eine Produkthaftung für Softwareprodukte MüKo/*Wagner*, § 2 ProdHaftG, Rn. 17 ff.; *Czychowski*, in: Fromm/ Nordemann, UrhG, § 69c Rn. 37f.; *Taeger*, CR 1996, 257, 258 ff.; *Lehmann*, NJW 1992, 1721, 1723 f.; *Cahn*, NJW 1996, 2899, 2904; *Günther*, Produkthaftung für Informationsgüter, 2001, S. 189 ff.; *Zech*, Information als Schutzgegenstand, 2012, S. 342 f.

39 *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 19: „intelligente, lernfähige Maschine zur Erweiterung der menschlichen Handlungsmöglichkeiten in der physikalischen Welt“ [wobei in dieser Definition auch schon Lernfähigkeit bzw. Autonomie als zusätzliche Komponente enthalten ist]; *Zech*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 165: „IT-gesteuerten Maschinen (Verknüpfung von Sensoren, Computersteuerung und Aktuatoren)“; *Yuan*, RW 2018, 477, 478: „Maschinen [...], die über Sensoren ihre physikalische Um-

Roboter entsprechend der klassischen Definition auch als Verbindung von Sensoren, Steuerung und Aktuatoren (sense, think [Informationsverarbeitung], act) definieren.[40] Bezogen auf den Automatisierungsbegriff (s.o. A.I.2) bedeutet Robotik Automatisierung von Hardware. Hier liegt auch der Schwerpunkt der Regelungen zu Fahrzeugen mit hoch- oder vollautomatisierter Fahrfunktion (s.o. A.V.2).

II. Digitale Systeme können unmittelbar in die physische Welt wirken

Wie bereits ausgeführt, ermöglicht es die Robotik informationsverarbeitenden Systemen, unmittelbar in der körperlichen Welt wirksam zu werden. Dies wurde früher noch als Ausnahmephänomen gesehen. Ein früher Fall, in dem ein Programmierfehler unmittelbare physische Schädigungen (des menschlichen Körpers) bewirkt hat, ist derjenige des medizinischen Bestrahlungsgerätes „Therac-25" aus den 1980er-Jahren.[41] Durch Robotik werden virtuelle Systeme physisch (s.o. A.IV).[42] Virtuelle Systeme stellen damit den Gegenbegriff zu Robotern dar, Informationseingabe und Informationsausgabe erfolgen bei ihnen nicht unmittelbar aus der realen Welt, sondern entweder über Menschen oder über Netzwerke. Dagegen erfolgt bei Robotern zumindest die Ausgabe zwingend über Aktuatoren. Die Eingabe kann variieren, erfolgt aber zumindest zu einem gewissen Teil über Sensoren des Roboters. *Bösl* beschreibt daher vier Phasen in der Entwicklung der Robotik von der automatisierten Steuerung über komplexere Steuerungen mit Sensoren zu mobilen und letztlich autonomen Robo-

gebung wahrnehmen, diese Daten mittels Prozessoren verarbeiten und über Aktuatoren auf ihre Umgebung physisch einwirken". Der Hardwareaspekt lässt sich als Teil der Maschinendefinition auffassen; vgl. die Definition einer Maschine nach Art. 2 Abs. 2 lit. a Richtlinie 2006/42/EG des Europäischen Parlaments und des Rates vom 17. Mai 2006 über Maschinen und zur Änderung der Richtlinie 95/16/EG (Neufassung), insbesondere die Grunddefinition im ersten Spiegelstrich: „eine mit einem anderen Antriebssystem als der unmittelbar eingesetzten menschlichen oder tierischen Kraft ausgestattete oder dafür vorgesehene Gesamtheit miteinander verbundener Teile oder Vorrichtungen, von denen mindestens eines bzw. eine beweglich ist und die für eine bestimmte Anwendung zusammengefügt sind". Allerdings wird gerade im Zusammenhang mit künstlicher Intelligenz auch von virtuellen Maschinen gesprochen, etwa bei den support vector machines (SVM), vgl. *Zech*, ZfPW 2019, 198, 201 m.w.N. Der Begriff robot wird ebenfalls auch für bestimmte Software-Agenten verwendet. Um diese rein virtuellen Systeme geht es aber bei der klassischen Roboter-Definition gerade nicht.

40 *Bekey*, Autonomous Robots, 2005, S. 2: „a machine that senses, thinks, and acts"; vgl. *Christaller* u.a., Robotik, 2001, S. 18 ff. (die auf die Definitionen in technischen Standards hinweisen); *Lohmann (Müller)*, AJP 2014, 595, 596; *Calo*, 70 Md. L. Rev. 571 (2011), 573.

41 Dazu *Leveson/Turner*, Computer 1993, 18 ff.; *Baase*, A Gift of Fire – Social, Legal, and Ethical Issues for Computing and the Internet, 3. Aufl. 2008, S. 425 ff.; *Zech*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 168.

42 *Calo*, 70 Md. L. Rev. 571 (2011), 593 ff.: „robots as physical PCs".

tern.[43] Mit Autonomie ist hier in erster Linie die Steuerung charakterisiert, womit die Lernfähigkeit der Informationsverarbeitung angesprochen ist (dazu unter C.). Parallel dazu haben sich aber auch die Freiheitsgrade der Hardware erhöht (Mobilität). Die Autonomie hochentwickelter Roboter ergibt sich aus dem Zusammenspiel von Mobilität und Lernfähigkeit.

III. Aktuelle Entwicklungen führen zu mehr physischen Freiheitsgraden

Zu einer höheren Mobilität haben diverse Weiterentwicklungen im Hardwarebereich, insbesondere aber verbesserte Energiespeicherung und leistungsfähigere IT (besseres Gewicht-Leitungs-Verhältnis), beigetragen. Als „technical drivers" der Entwicklung nennt *Pratt* exponentielles Wachstum der Computerleistung, Verbesserung der Geräte für die Herstellung (insbesondere 3D-Druck), Verbesserungen bei der Speicherung elektrischer Energie, Verbesserung der Energieeffizienz elektronischer Geräte, exponentielles Wachstum von Größe und Leistung des Internet, exponentielles Wachstum der weltweiten Datenspeicherung und exponentielles Wachstum der weltweiten Gesamt-Computerleistung.[44] Leistungsfähigere (leichtere) Sensoren und die Vernetzung von Hardware (insbesondere über Funknetze, dazu unter D.) kommen hinzu.

IV. Zunehmende Verbreitung in verschiedensten Lebensbereichen

Mehr Freiheitsgrade der Hardware (Mobilität) und der Steuerung (Lernfähigkeit) ermöglichen es, dass Roboter in immer komplexeren Umgebungen eingesetzt werden. Während klassische Industrieroboter noch ortsfest in einem abgeschirmten Raum arbeiteten, sind modernere Industrieroboter bereits beweglich und erste Roboter werden auch im freien Straßenverkehr eingesetzt. Bewegliche Haushaltsroboter, die einfachere Aufgaben erledigen können, sind bereits als Saugroboter Alltag. Zu erwähnen sind auch Drohnen. In der Pflege gibt es bereits Prototypen komplexerer beweglicher Roboter. Ziel vieler Projekte sind humanoide Roboter, die direkt in der menschlichen

43 *Bösl* nach *Krohn*, FAZ, 19.2.2016, 16; vgl. *Zech*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 165 f.

44 *Pratt*, Journal of Economic Perspectives 29 (2015), 51, 53 ff.

Umgebung wirken können.[45] Ihre Menschenähnlichkeit bedeutet ein Höchstmaß an Mobilität und Lernfähigkeit.

Wichtig ist, dass diese Möglichkeit des Einsatzes in immer komplexeren Umgebungen bedeutet, dass Roboter auch in immer mehr Lebensbereichen eingesetzt werden können. Daher steht zu erwarten, dass Roboter – wie virtuelle digitale Systeme – mehr oder weniger alle Lebensbereiche durchdringen werden. Der engere physische Kontakt mit Menschen wiederum bedingt ein höheres Potential für Schädigungen an Leib und Leben (dazu unter V.).

Mobilität und Einsatzbereiche stellen auch wichtige Kategorien für die Beurteilung von Robotern dar. Es macht – auch für die Risikobeurteilung – einen großen Unterschied, ob Roboter ortsfest oder beweglich sind, ob sie in geschlossenen Umgebungen oder im freien Verkehr eingesetzt werden, und, ob sie im unternehmerischen Bereich oder im privaten Bereich eingesetzt werden (wobei auch bei Unternehmen je nach Art und Größe Unterschiede in der Beherrschung der Hardware bestehen). Weitere Aspekte wie Lernfähigkeit und Vernetzung kommen noch hinzu (für eine Zusammenfassung aller relevanten Aspekte s.u. unter E.IV). Ähnlich vertritt auch *Lohmann* eine Kategorisierung von Robotern nach Lern- und Entscheidungsfähigkeit einerseits und Strukturiertheit des Einsatzbereichs andererseits.[46]

V. Risiken der Robotik

Die Robotik bringt spezifische Risiken, d.h. Möglichkeiten des Schadenseintritts,[47] mit sich: Neben der unmittelbaren physischen Schädigung durch maschinelle Informationsverarbeitung sind als besonderer Fall die Schädigung von Menschen bei Mensch-Maschine-Interaktionen und als zukünftiges Szenario die Schädigung durch Schwarmverhalten zu nennen. Das Risiko computergesteuerter Maschinen und das besondere Risiko ihrer Teilnahme am Verkehr lassen sich auch zusammenfassend als *Robotikrisiko* bezeichnen.[48] Dieses Robotikrisiko kann als Spezialfall des Automati-

45 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 139.

46 *Lohmann*, AJP 2017, 152 f.

47 Zum Risikobegriff s.u. Teil 2, A.II.

48 *Zech*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 172 ff.: „Roboterrisiken“ sind Komplexitätsrisiko (Risiko computergesteuerter Maschinen) und Mobilitätsrisiko (Risiko durch Teilnahme am allgemeinen Verkehr) sowie Vernetzungsrisiko (wird hier – wie auch das Autonomierisiko – getrennt behandelt).

sierungsrisikos, also des Risikos der Automatisierung von Vorgängen im Allgemeinen, gesehen werden.

1. Auftreten unmittelbarer physischer Schadensverursachung

Die Verbindung von digitalen Systemen mit Aktuatoren führt zu unmittelbaren physischen Schädigungswegen, die bislang bei digitalen Systemen eher die Ausnahme waren.[49] Verletzungen der körperlichen Rechtsgüter, Eigentum und körperliche Unversehrtheit durch digitale Systeme treten infolgedessen vermehrt auf (frühes Beispiel „Therac-25"). Ob dies zu einer Erhöhung des Risikos von Körper- oder Eigentumsschäden in einem bestimmten Bereich führt, richtet sich danach, wie hoch das Risiko der vorher eingesetzten Lösung war.

Virtuelle Systeme können allenfalls unkörperliche Rechtsgüter unmittelbar schädigen. Eine solche unmittelbare nicht-physische Schädigung kommt etwa bei Persönlichkeitsrechten[50] und Immaterialgüterrechten in Betracht (etwa Rufschädigung durch Verbreitung von Information im Internet). Eigentum[51] oder Leib und Leben können sie nur mittelbar schädigen, indem die ausgegebene Information für schädigende Handlungen Dritter kausal wird.

Infolge der Entwicklung der Robotik tritt deshalb neben das bekannte Problem der mittelbaren Schädigung durch digitale Systeme (fehlerhafte Software führt zu Computerausfall und mittelbar zu weiteren Schäden, fehlerhafte Software erzeugt fehlerhafte Information, Plattformen ermöglichen vorsätzliche Schädigungen) der unmittelbare physische Schädigungsweg.

Damit ist auch eine erste wichtige Kategorisierung digitaler Systeme angelegt, nämlich in solche Systeme, die unmittelbare physische Schäden verursachen (bzw. auf naturgesetzlichem Wege) können, und solche, die dies nur durch Informationsübermittlung bzw. -verbreitung (auf kommunikativem Wege, Informationsübermittlung an Menschen) vermögen.[52]

---

49 *Spindler*, in Hilgendorf (Hrsg.), Robotik im Kontext von Recht und Moral, 2014, S. 63, 69 ff.; *Borges*, in Lohsse/Schulze/Staudenmayer (Hrsg.), Liability for Artificial Intelligence and the Internet of Things, 2019, S. 145, 146 f.

50 Dazu *Oster*, UFITA 2018, 14 ff.

51 Zu der Frage, ob Eigentum auch unkörperliche Befugnisse zuweist *Zech*, AcP 219 (2019), 488, 564 ff.

52 Ähnlich *Riehm/Meier*, in DGRI Jahrbuch 2018 (erscheint demnächst), S. 5: physische Schäden einerseits, ehrverletzende Suchvorschläge und Beleidigungen andererseits.

Zu beachten ist jedoch, dass zunehmend komplexe Sensorik und Steuerung auch zu einer Verringerung des Risikos einer physischen Schädigung durch Roboter führen kann. So sind neuartige Roboter in der Lage, aktiv auf menschliche Bewegungen zu reagieren und dadurch Schädigungen zu vermeiden (Forschungsgebiet der Physical Human-Robot Interaction).[53]

2. Vermehrte physische Interaktion mit Menschen

Die zunehmende Verbreitung von Robotern im Alltag erhöht die Zahl der Interaktionen mit digitalen Hardware-Systemen und damit auch die Eintrittswahrscheinlichkeit entsprechender Schädigungen. Die Verbreitung von Robotern wird eine „neue Intensität der Mensch-Maschine-Interaktion“[54] bewirken, die sich nicht nur bei der Schadensverursachung, sondern auch in ihrer Nützlichkeit oder in Effekten wie der Anthropomorphisierung digitaler Systeme von der Interaktion mit virtuellen digitalen Systemen (über das Mobiltelefon etc.) unterscheidet.

Ähnlich weist auch *Teubner* auf das besondere „Verbundrisiko“ hin, das sich aus der engen Kooperation von Menschen und Softwareagenten ergibt.[55] Allerdings bezieht sich dieses Verbundrisiko auch auf rein virtuelle Systeme. Durch die physische Interaktion wird es verstärkt.

Es geht also nicht um die bloße Robotik, sondern darum, dass sich Roboter im Alltag bewegen. Dies wird durch die hohe Zahl physischer Freiheitsgrade ermöglicht. Die Teilnahme frei beweglicher Maschinen am allgemeinen Verkehr lässt sich als besonderes Risiko auffassen.[56]

---

53 *Haddadin/Croft*, in Siciliano/Khatib (Hrsg.), Springer Handbook of Robotics, 2. Aufl. 2016, Kap. 69 (Physical Human–Robot Interaction); *Haddadin*, Towards Safe Robots: Approaching Asimov’s 1st Law, 2011, http://darwin.bth.rwth-aachen.de/opus3/volltexte/2011/3826/pdf/3826.pdf (zuletzt aufgerufen am 10.10.2019); *Giannaccini*, Safe and Effective Physical Human-robot Interaction, 2015; *Barattini* u.a. (Hrsg.), Human-Robot Interaction, 2019.

54 *Yuan*, RW 2018, 477, 478.

55 *Teubner*, AcP 218 (2018), 155, 164: „Verbundrisiko, das auf die enge Kooperation von Mensch und Softwareagent zurückzuführen ist“; ebenso *Cornelius*, ZRP 2019, 8, 10.

56 *Zech*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 173 f.

3. Schwarmrobotik

Eine besondere Herausforderung stellt die Schwarmrobotik dar.[57] Sie zieht besondere Vorteile, aber auch Risiken, aus dem Zusammenwirken zahlreicher Roboter im Schwarm. Aus der Komplexität des Gesamtsystems (Schwarms) resultieren Effekte, die zu unvorhergesehenen Schädigungen führen können. Das Schwarmrisiko ist daher dem Vernetzungsrisiko vergleichbar (s.u. D.). Schwärme können spontanes emergentes Verhalten zeigen.

D. Lernfähigkeit (Autonomie): Trainieren statt Programmieren

Derzeit besonders intensiv diskutiert wird die Autonomie als neuartige Eigenschaft digitaler Systeme. Im Kern handelt es sich dabei um die Fähigkeit, zu lernen. Während entsprechende Systeme im Bereich der Robotik noch Zukunftsmusik sind (jedenfalls in der breiten Anwendung), finden sich virtuelle autonome Systeme bereits in alltäglichen Anwendungen wie Spracherkennung oder maschineller Übersetzung. Sie führen zu spezifischen Risiken, die man als Autonomierisiko[58] bezeichnen kann.

Zunächst soll das Grundprinzip dargestellt werden, der Übergang vom Programmieren zum Trainieren. Danach werden zwei Unterscheidungen vorgestellt, die für das Verständnis der Lernfähigkeit wichtig sind: zum einen, ob man die Regeln, nach denen die Informationsverarbeitung erfolgt, als Modell der Umwelt oder als Verhalten bzw. Verhaltensregeln versteht, zum anderen, ob sie explizit oder implizit repräsentiert werden. Daran anknüpfend soll dargestellt werden, welche Vorgaben durch den Programmierer oder durch den Trainer gemacht werden können. Schließlich werden verschiedene Möglichkeiten dargestellt, wie lernfähige digitale Systeme realisiert werden können. Insbesondere mehrschichtige neuronale Netze (deep learning) haben hier einige Prominenz erlangt. Abschließend wird untersucht, welche spezifischen Risiken lernfähige digitale Systeme mit sich bringen.

---

57 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 149 ff.

58 *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 175 f.; *ders*. ZfPW 2019, 198, 209; *Teubner*, AcP 218 (2018), 155, 164.

I. Grundprinzip: System lernt Regeln, nach denen die Informationsverarbeitung erfolgt, statt sie vorgegeben zu bekommen

Lernfähige Systeme können die Regeln, nach denen die Informationsverarbeitung erfolgt, „lernen", statt sie als Programm vorgegeben zu bekommen. Dies ermöglicht es, auch mit komplexen Informationen bzw. Umgebungen zurechtzukommen. Man spricht auch von Maschinenlernen (machine learning).[59]

1. Trainieren statt Programmieren (Regeln beruhen auf Gelerntem)

Das Grundprinzip lernfähiger digitaler Systeme lässt sich auf folgende Formel bringen: Sie werden nicht programmiert, sondern trainiert. Die Eingabe der Regeln, nach denen die Informationsverarbeitung erfolgt, und nicht etwa nur der zu verarbeitenden Information, wird dabei nicht durch den Menschen (Programmierer) vorgenommen, sondern eigenständig, d.h. durch das System, indem es aus Daten „lernt" (insbesondere über eigene Sensoren, aber auch durch Vernetzung). Das System erstellt oder modifiziert die Regeln, nach denen die Informationsverarbeitung erfolgt, als Reaktion auf die zu verarbeitende Information (Daten).

Dies entspricht auch der zentralen Definition für Autonomie digitaler Systeme (autonome Agenten), die sich etwa bei *Russell/Norvig* findet: „To the extent that an agent relies on the prior knowledge of its designer rather than on its own percepts, we say that the agent lacks autonomy."[60] Lernfähige Systeme können sich – zumindest in einem gewissen Maße – auf ihre eigene Wahrnehmung (its own percepts) statt auf das Wissen ihres Entwicklers (prior knowledge of its designer) verlassen. Dies erreichen sie, indem sie die Regeln ihrer Informationsverarbeitung aus den eigenen Wahrnehmungen ableiten und so lernen.

Damit lassen sich zwei grundlegende Arten der Regelgewinnung bzw. zwei verschiedene Grundprinzipien von KI unterscheiden: die modellbasierte Regelbestimmung

59 Zum Maschinenlernen (machine learning) *Ertel*, Grundkurs Künstliche Intelligenz, 4. Aufl. 2016, S. 191 ff.; *Sorge*, in Hornung (Hrsg.), Rechtsfragen der Industrie 4.0, 2018, S. 139, 140 ff.; *Alpaydin*, Machine Learning, 2016. Vgl. auch *Gausling*, DZ 2019, 335.

60 *Russell/Norvig*, Artificial Intelligence: A Modern Approach, 3. Aufl. 2009, 40. Vgl. Lohmann, *Reichwald/Pfisterer*, CR 2016, 208 (210 f.); *Wildhaber/Lohmann*, AJP 2017, 135 (136). Der Begriff der Autonomie wird unter VI. noch ausführlicher beleuchtet.

durch Programmierer (entsprechend der klassischen KI) und die selbsttätige Regelbestimmung durch Lernen bzw. Aufnahme von Information aus der Umwelt. Diese beiden Prinzipien können durchaus auch in hybriden Systemen kombiniert werden (Algorithmen werden auch als Bausteine von Maschinenlern-Systemen eingesetzt). Auch der Modellbegriff wird noch genauer dargestellt (unter II.). Da die Regelbestimmung durch Programmierer explizit durch symbolische Logik (codiert in Form einer Programmiersprache) erfolgen kann, wird die dargestellte grundlegende Unterscheidung häufig auch als symbolische Argumentation versus machine learning charakterisiert.[61]

2. Praktische Relevanz: Problem der zu großen Zustandsräume

Grund für den Erfolg lernfähiger digitaler Systeme (und für die Probleme klassischer KI-Ansätze) ist, dass komplexere Aufgaben zu viele mögliche Lösungen haben, als dass diese im Vorhinein vollständig durchdacht werden könnten. Während ein einfaches Spiel wie Tic-Tac-Toe weniger als 10^5 Spielverläufe kennt, sind es beim Schach bereits um die 10^{120} (und damit mehr, als es Atome im Universum gibt).[62] Daher ist kein Schachprogramm in der Lage, sämtliche möglichen Spielverläufe zu kalkulieren, sondern kann eine erfolgreiche Strategie nur mit heuristischen Methoden entwickeln. Es handelt sich um eine Planungsaufgabe mit zu vielen Zustandsräumen. Geht es darum, einen Roboter im Alltag zu bewegen, ist die Zahl möglicher Situationen nicht einmal mehr zuverlässig abschätzbar. Daher können humanoide Roboter oder auch selbstfahrende Fahrzeuge für den Straßenverkehr nicht vollständig programmiert werden.

61 *Kaplan*, Künstliche Intelligenz, 2015, S. 36 ff.; *Bermudez*, Cognitive Science, 2. Aufl. 2014, S. 138 ff.

62 Zu Tic-Tac-Toe und Schach *Kaplan*, Künstliche Intelligenz, 2015, S. 17. Die Zahl der möglichen Spielverläufe beim Schach übersteigt aber die Zahl der Atome im Universum, so dass ein heuristisches Verfahren gewählt werden muss.

II. Modell versus Verhalten: die Regeln, nach denen die Informationsverarbeitung erfolgt

Bisher wurde nur von Regeln, nach denen die Informationsverarbeitung erfolgt, gesprochen. Hier gibt es aber zwei verschiedene Auffassungen von Lernen, die sich auch in der Herangehensweise bei digitalen Systemen spiegeln: Lernen durch Modellbildung bzw. -anpassung und Lernen durch Generalisierung.

1. Lernen durch Modellbildung und Lernen durch Generalisierung

Grundsätzlich lassen sich zwei verschiedene Arten des Lernens unterscheiden, nämlich entweder durch Verallgemeinerung von Erfahrungen (experience generalisation models) oder durch Veränderung der Repräsentation eines Problemfelds (models transforming a representation of a problem domain).[63] Entweder das System lernt durch Generalisierung von Input, oder es lernt durch Bildung eines inneren Modells der Umgebung (Ontologie).

Die ideengeschichtlich ältere Möglichkeit, zu lernen, besteht darin, Repräsentationen eines Problemfelds (Problemdomäne) zu modifizieren. Das System muss also über ein Modell des Bereichs der realen oder virtuellen Welt, in dem es angewendet werden soll, verfügen. Dies basiert letzten Endes auf der Idee der klassischen Kognitionswissenschaft, die davon ausgeht, dass ein informationsverarbeitendes System, um Probleme zu lösen, ein vollständiges Modell des Problemlösungsfeldes besitzen muss. Die Regeln, nach denen die Informationsverarbeitung abläuft, ergeben sich aus diesem Modell bzw. können mit diesem gleichgesetzt werden. Die klassische KI scheiterte gerade an dem Versuch, vollständige symbolische Repräsentationen komplexer Problemfelder zu erstellen. Eine eigenständige Bildung symbolisch repräsentierter Modelle durch künstliche Systeme ist nicht möglich, wohl aber die Anpassung vorgegebener Modelle.[64]

Die andere Art, zu lernen, besteht darin, Erfahrungen zu generalisieren. Hier lernt das System, indem es Muster in den Erfahrungen erkennt. So formuliert *Kaplan*: „Computerprogramme lernen, indem sie Muster aus Daten extrahieren.“[65] Da Lernen durch

63 *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 230 ff.

64 *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 231.

65 *Kaplan*, Künstliche Intelligenz, 2015, S. 44.

Generalisierung keine symbolische Repräsentation des Problemfeldes voraussetzt, die bereits vorgegeben werden muss, lässt es sich erheblich einfacher realisieren. In der aktuellen Diskussion um Maschinenlernen steht daher Lernen durch Generalisierung im Vordergrund bzw. teilweise wird nur die Generalisierung als Grundlage maschinellen Lernens behandelt.[66]

Ob sich in neuronalen Netzen ein implizites Modell des Problemfeldes ergibt oder ob diese nur generalisieren, ist letzten Endes eine philosophische Frage. Interessant ist, dass menschliche Gehirne durchaus als Modelle ihrer Umgebung verstanden werden können, einerseits implizit, andererseits aber auch explizit als Summe der Erfahrungssätze, die laufend angepasst werden können. Epistemologisch ähnelt diese sehr dem Web of Belief (Netz von Überzeugungen), das dem Wissens- bzw. Erkenntniskonzept des Holismus zugrunde liegt.[67] Künstliche Netze sind dagegen (noch) nicht imstande, eigenständig eine symbolische Repräsentation ihres Erfahrungsraumes zu bilden.

Durchaus lässt sich aber durch Generalisierung Gelerntes als implizites Modell auffassen. Nur stellt dieses kein vollständiges Modell des Problemfeldes dar, was wiederum unproblematisch erscheint, weil zumindest Modelle der Realität ohnehin immer nur Vereinfachungen sein können. Modellbildung und Generalisierung widersprechen sich also nicht notwendig. Zusammenfassend lässt sich festhalten: Den Regeln, nach denen die Informationsverarbeitung erfolgt, liegt ein explizites oder implizites Modell zugrunde (bei der Programmierung bereits beim Programmierer), und die Regeln führen zu einem bestimmten Verhalten (das durch die Regeln beschrieben werden kann, wenn sie explizit sind).

2. Funktionalistischer, konnektionistischer und handlungsorientierter Ansatz

Auf die besondere Bedeutung des Verhaltens eines Systems geht die Unterscheidung zwischen Funktionalismus, Konnektionismus und Handlungsorientierung als Ansätze für die Gestaltung der Informationsverarbeitung ein, die *Mainzer* vorstellt.[68] Sie beruht auf der bereits vorgestellten Unterscheidung von Lernen durch Modellbildung und durch Generalisierung.

---

66 Etwa bei *Ertel*, Grundkurs Künstliche Intelligenz, 4. Aufl. 2016, S. 192.

67 *Quine/Ullian*, The Web of Belief, 1978.

68 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 142 ff.

Der funktionalistische Ansatz geht von einer vollständigen expliziten Repräsentation der Außenwelt im System aus: „Die Grundannahme des Funktionalismus besteht darin, dass es in Lebewesen wie in entsprechenden Robotern eine interne kognitive Struktur gibt, die Objekte der externen Außenwelt mit ihren Eigenschaften, Relationen und Funktionen untereinander über Symbole repräsentiert.“[69] Dieser Ansatz liegt insbesondere der klassischen Programmierung zugrunde. Voraussetzung dafür ist, dass bereits der Programmierer über ein solches Modell bzw. eine solche vollständige Repräsentation verfügt. Dies führt zu den genannten praktischen Schwierigkeiten der klassischen KI.[70]

Demgegenüber basiert der konnektionistische Ansatz auf der Wechselwirkung zwischen Einheiten eines komplexen Netzwerks als Grundlage für die Informationsverarbeitung. Dieser Ansatz liegt neuronalen Netzen zugrunde. Die Regeln, nach denen die Informationsverarbeitung abläuft, liegen nicht als Modell und auch nicht als symbolische Verhaltensregel vor (zumindest grundsätzlich, zu explainable AI s.u. III.). *Mainzer* formuliert das so: „Der konnektionistische Ansatz betont [...], dass Bedeutung nicht von Symbolen getragen wird, sondern sich in der Wechselwirkung zwischen verschiedenen kommunizierenden Einheiten eines komplexen Netzwerks ergibt. Diese Herausbildung bzw. Emergenz von Bedeutung und Handlungsmustern wird durch die sich selbst organisierende Dynamik von neuronalen Netzwerken [...] möglich.“[71]

Der dritte, handlungsorientierte Ansatz stellt auf das Verhalten des Systems ab. Hier ist die Einbettung in die Umwelt entscheidend, es kommt auf das Verhalten an. Dieser Ansatz ist vor allem für Roboter wichtig, da es hier auf die Interaktion mit der Umwelt ankommt. Es geht um „verhaltensgesteuerte Artefakte [...], die sich an veränderte Umweltbedingungen anzupassen vermögen“.[72]

---

69 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 142 f.

70 *Ertel*, Grundkurs Künstliche Intelligenz, 4. Aufl. 2016, S. 191: „Die Forderung nach maschinellen Lernverfahren ergibt sich aber auch aus dem Blickwinkel des Software-Entwicklers, der zum Beispiel das Verhalten eines autonomen Roboters programmieren soll. Die Struktur des intelligenten Verhaltens kann hierbei so komplex werden, dass es auch mit modernen Hochsprachen wie Prolog oder Python sehr schwierig oder sogar unmöglich wird, dieses annähernd optimal zu programmieren. Ähnlich wie wir Menschen lernen, werden auch heute schon bei der Programmierung von Robotern maschinelle Lernverfahren eingesetzt [...], oft auch in einer hybriden Mischung aus programmiertem und gelerntem Verhalten.“

71 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 144.

72 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 144.

3. Praktische Relevanz hybrider Systeme

Mainzer selbst betont, dass gerade die Kombination der drei Ansätze praktische Erfolge ermöglicht und dass sich auch in der menschlichen Intelligenz alle drei Aspekte finden.[73] Spektakuläre Erfolge im Bereich der künstlichen Intelligenz wurden durch die Verbindung verschiedener Systeme möglich: So handelt es sich bei selbstfahrenden Fahrzeugen um die hierarchisch angeordnete Verbindung unterschiedlicher Systeme von der Mustererkennung auf niedriger Ebene bis zur Planung auf höchster Ebene. Auch AlphaGo ist eine Kombination verschiedener Systeme, insbesondere von klassischer algorithmischer Problemlösung und Mustererkennung durch deep learning. Solche hierarchischen Anordnungen verschiedener Systeme sind auch aus wirtschaftlicher und haftungsrechtlicher Sicht interessant, da Subsysteme auch von Zulieferern entwickelt und/oder trainiert werden können.

III. Explizit oder implizit: Repräsentation des Gelernten

Eine weitere wichtige Unterscheidung, die bereits unter II. angesprochen wurde, ist die zwischen expliziter und impliziter Repräsentation, die nicht nur die zu verarbeitende Information, sondern auch die Regeln, nach denen die Informationsverarbeitung ausgeführt wird, betrifft.

1. Symbolische Repräsentation versus implizite Repräsentation (embodied cognition)

Grundsätzlich können die Regeln, nach denen sich das Verhalten eines digitalen Systems bestimmt, entweder symbolisch repräsentiert sein, oder nur im Zustand des Systems repräsentiert sein, ohne dass eine symbolische Codierung verwendet wird (embodied cognition). Mangels Codierung gibt es dabei auch keine Bedeutungsebene mehr, die von einem menschlichen Betrachter „gelesen“ werden könnte. Dies bedeutet, dass das Verhalten des Systems nicht mehr durch Untersuchung des zugrundeliegenden Programms überprüft werden kann (dazu unter 3.).

---

73 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 144 f; vgl. das Zitat von *Ertel* in Fn. 79.

2. Explainable AI

Ein wichtiges Forschungsfeld beschäftigt sich gerade damit, embodied cognition wieder verständlich zu machen. Unter dem Stichwort Explainable AI (XAI) geht es darum, Implizites wieder explizit machen.[74] Im Hinblick auf neuronale Netze wird davon gesprochen, die „black box" des trainierten Systems verständlich zu machen.[75] Dies erlaubt nicht nur eine Erklärung des Gelernten (eine Beschreibung ließe sich hier auch durch Beobachtung des Verhaltens erzielen, s.u. 3.), sondern insbesondere auch des Lernvorgangs. Damit wird eine Zuordnung des erlernten Verhaltens zu bestimmten Inputdaten möglich, auch wenn diese nicht vollkommen eindeutig sein kann (wie bei Programmierungen), sondern nur mit einer gewissen Wahrscheinlichkeit angegeben werden kann.

3. Bedeutung für die Beurteilung digitaler Systeme

Trainieren statt Programmieren stellt auch das Sicherheitsrecht vor neue Herausforderungen. Die Sicherheitsbeurteilung muss nach anderen Prinzipien erfolgen, da die Steuerung nicht mehr auf einem feststehenden Steuerungscode beruht, der direkt untersucht werden kann. Grundsätzlich gibt es aber zwei verschiedene Methoden der Sicherheitsbeurteilung: Statt den Code zu überprüfen, kann man auch Testläufe durchführen. Entsprechend gibt es sog. Blackbox- und Whitebox-Tests.[76] Welcher Grad von Sicherheit verlangt wird und welche Wege, diese zu erreichen, akzeptabel sind, ist eine normative Frage. Hier wird vor allem der technischen Normierung eine große Bedeutung zukommen.[77]

---

74 https://gi.de/informatiklexikon/explainable-ai-ex-ai/ (zuletzt aufgerufen am 10.10.2019).

75 *Castelvecchi*, 538 Nature 21 (2016); *Voosen*, How AI detectives are cracking open the black box of deep learning, http://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning; *Knight*, The Dark Secret at the Heart of AI, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (beide zuletzt aufgerufen am 10.10.2019).

76 *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, S. 44 ff.

77 Dazu Deutsches Institut für Normung (DIN), Warum auch die KI Normen braucht, https://www.din.de/de/din-und-seine-partner/publikationen/din-magazin/warum-auch-die-ki-normen-braucht-320570 (zuletzt aufgerufen am 10.10.2019).

IV. Vorgaben durch den Programmierer

Aus haftungsrechtlicher Sicht ist vor allem interessant, welchen Einfluss die verschiedenen Akteure auf das Verhalten des lernfähigen digitalen Systems haben, das letztendlich zu einem Schaden führt. Dabei ist zunächst der Programmierer in den Blick zu nehmen, der bei der klassischen Programmierung die volle Kontrolle über das Verhalten des Systems hat.

Diese Kontrolle wird bei lernfähigen Systemen geringer, dementsprechend wird auch die Vorhersehbarkeit des Verhaltens für den Programmierer geringer (dazu unter VII.). Damit beruht das System nicht mehr (oder zumindest in geringerem Maße) auf Wissen des Programmierers, sondern erwirbt das erforderliche Wissen erst. Mit *Specht/Herold* kommt es zur „Generierung von Wissen durch Erfahrungen der Maschine selbst".[78] Dies gilt nicht nur für Systeme, die Wissen repräsentieren, sondern in Form von Verhaltensregeln für alle lernfähigen Systeme.[79]

Das Wissen des Programmierers kann also kleiner sein als bei klassischer Programmierung, dennoch hat er Möglichkeiten der Einflussnahme auf den Lernprozess und damit auf das Lernergebnis.

1. Bestimmung der Lernfähigkeit

Zum einen hat es der Programmierer in der Hand, digitale Systeme überhaupt lernfähig auszugestalten. Auch kann er bestimmen, welche Teile eines Steuerungscodes angepasst werden können bzw. ganz allgemein kann er sich für oder gegen die Verwendung einer bestimmten Architektur entscheiden.

2. Setzen von Grenzen: Garantien und Spezifikationen des Verhaltens

Zum anderen gibt es die Möglichkeit, durch Programmierung bestimmte abstrakte Grenzen des Lernbaren zu setzen. Die Möglichkeit, solche high-level-Spezifikationen zu setzen, ist ein wichtiger Forschungsgegenstand. Grenzen zu setzen, die auch theo-

78 *Specht/Herold*, MMR 2018, 40, 41.

79 Die epistemologische Frage, was unter Wissen zu verstehen ist, soll hier ausgeblendet bleiben. Die Unterscheidung von Modellen und Generalisierungen wurde bereits unter II. angesprochen.

retisch beweisbar sind (wie bei einer feststehenden Programmierung), sog. beweisbare Garantien, ist schwierig.[80]

Neben der softwaremäßigen Begrenzung lernfähiger Systeme, lassen sich lernfähige Roboter auch physisch abgrenzen (bei virtuellen Systemen wäre ein Abschneiden von Netzen erforderlich). Dies wurde bereits bei der Robotik angesprochen. Ganz allgemein können Systeme physisch und informationell isoliert werden, um sie so gefahrlos zu testen (Idee des Sandkastens). Bei der Anwendung hängt der Nutzen digitaler Systeme jedoch gerade damit zusammen, dass sie nicht isoliert betrieben werden müssen.

V. Vorgaben durch den Trainer

Als neuer Akteur, der ebenfalls Einfluss auf das Verhalten des digitalen Systems hat, kommt der Trainer hinzu. Je nach Trainingsverfahren hat er mehr oder weniger Kontrolle über das, was das System lernt.

1. Überwachtes Lernen, unüberwachtes Lernen und Lernen durch Verstärkung

Eine wichtige Unterscheidung, die sich aus der Art der verwendeten Systeme ergibt (also vom Programmierer angelegt ist), ist diejenige zwischen überwachtem Lernen, unüberwachtem Lernen und Lernen durch Verstärkung.[81]

Beim überwachten Lernen (supervised learning) ist der Prototyp des zu Lernenden bereits bekannt, d.h. der Trainer kennt die zu lernenden Regeln und wählt die Trainingsdaten entsprechend aus bzw. sorgt dafür, dass die Regeln in den Trainingsdaten implizit enthalten sind (etwa Bilder von Katzen und Bilder von Hunden mit entsprechender Kennzeichnung).

Beim verstärkenden Lernen (reinforcement learning) kennt der Trainer zwar die Regeln, muss aber nicht dafür sorgen, dass sie bereits in den Trainingsdaten enthalten sind. Erst durch die Verstärkung kann das lernfähige System aus den Trainingsdaten

80 Für den Hinweis auf high-level-Spezifikationen und beweisbare Garantien danke ich *Malte Helmert*.

81 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2016, S. 115 ff.; *Kelleher/Tierney*, Data Science, 2018, S. 99 f.; *Ng/Soo*, Data Science – Was ist das eigentlich?, 2018, S. 8 ff.; *Sorge*, in Hornung (Hrsg.), Rechtsfragen der Industrie 4.0, 2018, S. 139, 140 f.; *Yuan*, RW 2018, 477, 485 ff.

Regeln extrahieren (Beispiel wieder Bilder von Katzen und Hunden, jedoch ohne entsprechende Kennzeichnung – als Verstärkung meldet der Trainer an das System, ob es jeweils zutreffend entschieden hat).
Beim unüberwachten Lernen werden dem System die geringsten Vorgaben gemacht. Das System soll dabei aus den Trainingsdaten Regeln extrahieren, die auch dem Trainer noch nicht bekannt sind. Big-Data-Analysen, bei denen in Datensätzen bislang unbekannte Korrelationen aufgefunden werden sollen, lassen sich als unüberwachtes Lernen charakterisieren.

2. Datenquelle

Ganz allgemein kann über die Auswahl der Trainingsdaten Einfluss auf das Lernergebnis genommen werden (was auch die Möglichkeit des Missbrauchs eröffnet). Häufig wird dies durch Bestimmung der Datenquelle geschehen. Hier zeigt sich auch ein Unterschied zwischen einfachen Betreibern und Nutzern, die keinen Einfluss auf die Trainingsdaten nehmen, und solchen, die auch als Trainer agieren.
Probleme mit der Qualität des Gelernten können sich aus „schlechten“ Trainingsdaten ergeben. Fehltrainings (etwa, dass Systeme bei der Bilderkennung versehentlich auf die Bildunterschriften trainiert werden) gehören zu den klassischen Problemen des Maschinenlernens. Anschaulich kann man auch vom Problem des „Garbage in, garbage out“ sprechen.

3. Beendigung des Lernprozesses

Eine letzte wichtige Einflussmöglichkeit (die auch im System angelegt sein muss) ergibt sich daraus, dass die Lernfähigkeit bei Maschinen auch zu einem bestimmten Zeitpunkt beendet werden kann. Damit ergibt sich die Möglichkeit, Systeme während einer Lernphase zu trainieren und nach Abschluss der Lernphase „einzufrieren“. Das Lernergebnis kann, wie bereits erwähnt, durch Vervielfältigung auch auf andere gleichartige Systeme übertragen werden. Jedenfalls eröffnet das Einfrieren des Lernergebnisses für Hersteller oder Bediener die Möglichkeit einer stärkeren Kontrolle.

VI. Lernfähigkeit als Autonomie

1. Autonomie als Unabhängigkeit von Vorgaben

Autonomie wurde bereits gleichgesetzt mit Lernfähigkeit. Dies ist allerdings nicht selbstverständlich. Wörtlich übersetzt bedeutet Autonomie, dass sich jemand oder etwas Regeln selbst setzt. Diese Regeln (νόμοι: Gesetze) sind bei digitalen Systemen die ausführlich dargestellten Regeln, nach denen die Informationsverarbeitung abläuft, die man auch als Verhaltensregeln beschreiben kann.

Lernfähigkeit bedeutet, dass diese Regeln zu einem geringeren Maße vom Programmierer bestimmt werden. Wie gezeigt hat aber auch der Trainer noch einen mehr oder weniger großen Einfluss auf das, was das System lernt. Nur den verbleibenden Rest an Eigenständigkeit bei der Regelgenerierung kann man wirklich als Autonomie bezeichnen. Dies wären dann nur spontane Effekte, die bei komplexen lernfähigen Systemen auftreten können.

Aus technischer (und haftungsrechtlicher) Sicht ist diese Differenzierung aber weniger wichtig. Es genügt, auf die Lernfähigkeit als besondere Eigenschaft abzustellen und den Einfluss verschiedener Akteure auf das Verhalten des Systems und damit auf durch das System verursachte Schäden (oder Nutzen) zu klären.

Der Begriff Autonomie wird zudem sehr vielfältig verwendet, zum einen im Bereich der Informationstechnologie selbst, darüber hinaus gibt es aber auch weitere Autonomiebegriffe in anderen Bereichen (z.B. Philosophie, Politik, Recht). Im Zusammenhang mit der haftungsrechtlichen Diskussion geht es um Autonomie als technischen Begriff, nicht um den Begriff aus der Rechtswissenschaft, Philosophie oder Sozialwissenschaften.[82]

---

82 In der rechtlichen Literatur finden sich u.a. folgende Definitionen: *John*, Haftung für künstliche Intelligenz, 2007, S. 20: „Fähigkeit eines Systems, aufgrund innerer Gesetzmäßigkeiten, mit und ohne äußere Einflüsse, über die Art und den Ablauf einer Handlung entscheiden zu können"; *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 226: „Softwaresysteme, die in einem gewissen Umfang intelligentes Verhalten aufweisen, zielorientiert handeln und über sog. ‚soziale' Fähigkeiten sowie Lernfähigkeit verfügen. [...] Ihre KI-spezifischen Eigenschaften können dazu führen, dass sie im Lauf der Zeit ihre Funktionalität und damit ihr Verhalten verändern, möglicherweise auch über die ursprünglich von Auftraggebern, Entwicklern und Anwendern intendierte Funktionalität hinaus. Sie besitzen damit ein gewisses Maß an Eigenständigkeit (aus Sicht von Entwicklern und Nutzern: Nicht-Determiniertheit), welches sie grundlegend von automatisierten Systemen wie etwa konventioneller Software und Industrierobotern unterscheidet."; *Riehm*, ITRB 2014, 113: „Gemeinsam ist den geschilderten Beispielen, dass technische Geräte – konkret deren Steuerungssoftware – auf immer weniger konkrete Benutzereingriffe angewiesen sind und immer komplexere Entscheidungen ‚selbst' treffen. Von außen entsteht der Eindruck der Unberechenbarkeit und Indeterminanz des ‚Verhaltens'."; *Bräutigam/Klindt*, NJW 2015, 1137: „Die Entwickler solcher künstlicher Intelligenz legen das Softwareverhalten nur indirekt in Form von Entwurfsentscheidungen fest. Eine spezifische Funktionalität oder vorherzusehende Individualität besitzen solche

Ausführliche monografische Diskussionen des Autonomiebegriffs aus rechtlicher Sicht (Autonomie bei technischen Systemen, Autonomiegrade) liegen bereits vor.[83] Zusammenfassend lassen sich zwei bzw. drei Definitionsstränge erkennen: Teilweise wird Autonomie als Unabhängigkeit von äußeren Einflüssen verstanden (autonomes physikalisches System[84]), teilweise als Fähigkeit zur Anpassung des eigenen Verhal-

Systeme nicht."; *Kirn/Müller-Hengstenberg*, Rechtliche Risiken autonomer und vernetzter Systeme, 2016, S. 125: „Möglichkeit zu selbstbestimmtem Verhalten, frei von externem Einfluss und externer Kontrolle [...] Agenten verfügen über ein Mindestmaß an Autonomie und besitzen eigene Ziele, die auch in Konflikt mit den Zielen des Agentenentwicklers und des Agentenanwenders stehen oder im Lauf der Zeit (z.B. durch Lernprozesse) geraten können."; *Reichwald/Pfisterer*, CR 2016, 208, 210 (Autonomiegrade): „Der Autonomiegrad ist demnach umso höher, je mehr das Verhalten von *intrinsischen Mechanismen* gesteuert wird und je mehr sich das System an Problemsituationen bzw. die Gesamtsituation eigenständig anpassen kann."; *Sosnitza*, CR 2016, 764, 765: „zeichnen sich autonome Systeme dadurch aus, dass sie in einem gewissen Umfang ihr eigenes Verhalten kontrollieren und ohne den Eingriff des Menschen handeln können"; *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 170: „Dies bedeutet, dass ein informationstechnisches System sein Verhalten auch ohne unmittelbaren Anlass von außen verändern kann."; *ders*., ZfPW 2019, 198, 200: „Autonomie bedeutet damit, dass sich ein Agent, beziehungsweise ein System auf eigene Wahrnehmungen verlässt, statt auf Eingaben des Anwenders angewiesen zu sein. [...] Ein selbstlernendes System reagiert also nicht nur auf Reize von außen, sondern passt als Reaktion auch sein Verhaltensmuster an (eventuell kann es dieses auch ohne Reize von außen anpassen, jedenfalls ohne menschliche Intervention). [...] Im Folgenden soll Autonomie mit der Fähigkeit des Selbstlernens gleichgesetzt werden."; *Borges*, NJW 2018, 977,978: „Der Begriff ‚Autonomie' eines Systems im Sinne einer funktionalen Einheit von Soft- und Hardware wird teilweise als die Fähigkeit beschrieben, Entscheidungen zu treffen und diese in der äußeren Welt unabhängig von externer Steuerung oder Einflussnahme umzusetzen. Autonome Systeme können daher als solche definiert werden, deren Verhalten nicht vollständig vorherbestimmt oder vorhersehbar ist."; *Hacker*, RW 2018, 243, 251 ff. (Grade von Autonomie): „Besitzt der KIA eigenständige, quasi-kognitive Fähigkeiten, die ein selbständiges Navigieren und die Durchführung im Einzelnen nicht vorhersehbarer, adaptiver Aktionen ermöglichen, so ist die Grenze vom bloßen Werkzeug zu einer personenähnlichen Entität überschritten." (251) „Bei vollständig autonomen KIA muss keine menschliche Aufsicht mehr in der konkreten Situation geleistet werden; der Agent ist hier in seinen Aktionen von seiner Umgebung unabhängig." (252) „Schwach autonome KIA hingegen bedürfen der ständigen menschlichen Anleitung und Überprüfung, da ihr System noch nicht im ausreichenden Maße auf neue Situationen reagieren kann." (253); *Specht/Herold*, MMR 2018, 40 f.: „Automatisierte Systeme folgen vorgegebenen Regeln, die durch den Menschen gesetzt werden. [...] Autonom agierende Systeme hingegen kommen ohne konkrete Voreinstellungen aus. [...] Gemeint ist die Generierung von Wissen durch Erfahrungen der Maschine selbst."; *Yuan*, RW 2018, 477, 481: „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"; *Cornelius*, ZRP 2019, 8 f.: „Die intelligenten Systeme träfen ihre Entscheidungen aufgrund eigener Erfahrungs- und Lernprozesse, könnten die ihrem Verhalten zugrunde liegenden Algorithmen selbständig ändern und bestimmten damit den Weg zur Zielerreichung selbst." Dazu auch *Sorge*, Softwareagenten, 2006, S. 8; *Sester/Nitschke*, CR 2004, 548, 549.

83 *Schulz*, Verantwortlichkeit bei autonom agierenden Systemen, 2015, S. 43 ff.; *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 29 ff.

84 *Christaller u.a.*, Robotik, 2001, S. 36: „Autonom ist ein gegebenes, technisches System dann, wenn es alleine auf Grund seiner inneren Zustände […] auch in Zukunft […] vollständig beschreibbar ist."

tens (autonome Agenten[85]), teilweise wird dies verbunden mit einem relevanten Einfluss auf die reale Welt.[86]

Im deutschen Sprachraum hat sich der Begriff Autonomie auch für Vollautomatisierung bei selbstfahrenden Fahrzeugen durchgesetzt (autonomes Fahren, basierend auf der Einteilung der Automatisierungsgrade durch das BASt[87]). Auch sonst wird Autonomie teilweise als vollständige Automatisierung (Grad 10) aufgefasst.[88] Diese Gleichsetzung von vollständiger Automatisierung und Autonomie entspricht der Definition als Unabhängigkeit von äußeren Einflüssen.

Die Definition des Europäischen Parlaments stellt auf Lernfähigkeit, eigenständiges Entscheiden und Interaktion mit der Umwelt ab.[89] Hier spielt also, wie überhaupt schwerpunktmäßig bei der Diskussion um Roboter, der Einfluss auf die reale Welt eine große Rolle.

Da der Begriff Autonomie insgesamt in der rechtlich-technischen Diskussion nicht klar definiert ist, erscheint er weniger geeignet zur Beschreibung des technischen Hintergrunds als der Begriff der Lernfähigkeit (welche der Fähigkeit zur Anpassung des eigenen Verhaltens entspricht).[90] In der amerikanischen Literatur findet sich auch der Begriff der Emergenz (emergence),[91] der jedoch vor allem für das Verhalten neuronaler Netze als emergente Eigenschaft dieser Netze passt, weniger für das Verhalten lernfähiger Systeme insgesamt.

---

85 *Russell/Norvig*, Artificial Intelligence: A Modern Approach, 3. Aufl. 2009, S. 40.

86 *Stiemerling*, CR 2015, 762, 765. Einfluss kann nicht-physisch sein wie etwa bei Trading-Systemen.

87 Siehe Fn. 41.

88 *Schulz*, Verantwortlichkeit bei autonom agierenden Systemen, 2015, S. 45.

89 Europäisches Parlament, P8_TA(2017)0051, Zivilrechtliche Regelungen im Bereich Robotik, Entschließung des Europäischen Parlaments vom 16. Februar 2017 mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL)), Erwägungsgründe Z: „dass die heutigen Roboter dank der beeindruckenden technischen Fortschritte des letzten Jahrzehnts nicht nur in der Lage sind, Tätigkeiten auszuüben, die früher einmal typische menschliche Tätigkeiten waren, welche ausschließlich den Menschen vorbehalten waren, sondern durch die Entwicklung bestimmter autonomer und kognitiver Merkmale – beispielsweise der Fähigkeit, aus Erfahrung zu lernen und quasi-unabhängige Entscheidungen zu treffen – den Akteuren, die mit ihrer Umwelt interagieren und diese ganz erheblich verändern können, immer ähnlicher geworden sind“; ebd., Erwägungsgründe Z.AA: „dass die Autonomie eines Roboters als die Fähigkeit definiert werden kann, Entscheidungen zu treffen und diese in der äußeren Welt unabhängig von externer Steuerung oder Einflussnahme umzusetzen, und in der Erwägung, dass diese Autonomie rein technologischer Art ist und ihr Grad davon abhängt, wie ausgeklügelt die Interaktion des Roboters mit seiner jeweiligen Umwelt konzipiert worden ist“. Dazu *Lohmann*, ZRP 2017, 168, 169.

90 Vgl. *Reichwald/Pfisterer*, CR 2016, 208, 210 f.; *Wildhaber/Lohmann*, AJP 2017, 135, 136; *Zech*, ZfPW 2019, 198, 200.

91 *Calo*, 103 Cal. L. Rev. 513 (2015), 538 ff.; *Balkin*, 6 Cal. L. Rev. Cir. 45 (2015), 51 f.; vgl. *Schirmer*, JZ 2019, 711.

2. Grade zunehmender Autonomie von digitalen Systemen (the extent that an agent relies on its own percepts)

Autonomie ist kein binäres Kriterium, sondern kann mehr oder weniger stark ausgeprägt sein.[92] Der Grad der Autonomie wird bestimmt durch Vorgaben des Programmierers (III.) und auch des Trainers (IV.), d.h. Vorgaben bedeuten zugleich Kontrolle bzw. Risikobeherrschung. Hohe Autonomie bedeutet hohe Flexibilität. So resümieren *Reichwald/Pfisterer*: „Der Autonomiegrad ist demnach umso höher, je mehr das Verhalten von *intrinsischen Mechanismen* gesteuert wird und je mehr sich das System an Problemsituationen bzw. die Gesamtsituation eigenständig anpassen kann."[93] Dies macht den praktischen Nutzen lernfähiger Systeme aus.

3. Autonomie und Kontrollmöglichkeiten

Wichtig erscheint aus haftungsrechtlicher Sicht, dass Autonomie verminderte Kontrollmöglichkeiten bedeutet. Entscheidende Aspekte sind dabei nach *Boden*[94] bzw. *Reichwald/Pfister*[95] der Anteil gelernten Wissens (statt modellhaft programmierten Wissens), der Anteil der Kontrolle über Eingangsparameter (eigenständige Problemlösung) und der Anteil ausgeübter Selbstkontrolle (eigenständige Problemdefinition). Diese kann wie dargestellt sowohl durch den Programmierer als auch durch den Trainer erfolgen. Systeme, die eigenständige Problemlösungen finden oder sich gar selbst ihre Ziele vorgeben, unterliegen auch einer geringeren Kontrolle (bereits die passende Vorgabe der Ziele durch die Entwickler ist nicht einfach).

---

92 *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 29 ff.; *Reichwald/Pfisterer*, CR 2016, 208, 210. Siehe bereits *Boden*, in dies. (Hrsg.), The Philosophy of Artificial Life, 1996, S. 95, 102. Fraglich ist, ob es eine maximale oder vollständige Autonomie geben kann (sofern man diese nicht mit Automatisierung gleichsetzt). *Horner/Kaulartz*, CR 2016, 7, 13 f., verwenden den Begriff des „vollautonomen" Systems.

93 *Reichwald/Pfisterer*, CR 2016, 208, 210.

94 *Boden*, in dies. (Hrsg.), The Philosophy of Artificial Life, 1996, S. 95, 102: „Three aspects of behavior – or rather, of its control – are crucial. First, the extent to which response to the environment is direct (determined only by the present state in the external world) or indirect (mediated by inner mechanisms partly dependent on the creature`s previous history). Second, the extent to which the controlling mechanisms were self-generated rather than externally imposed. And third, the extent to which inner directing mechanisms can be reflected upon, and/or selectively modified in the light of general interests or the particularities of the current problem in its environmental context."

95 *Reichwald/Pfisterer*, CR 2016, 208, 210.

VII. Realisierung durch verschiedene Architekturen: lernfähige Algorithmen, neuronale Netze, hybride Systeme

Die technische Realisierung lernfähiger Systeme soll nur kurz angeschnitten werden. Neben lernfähigen Algorithmen sind hier auch neuronale Netze zu nennen.

1. Lernfähige Algorithmen

Algorithmen können so gestaltet werden, dass sie durch lernen modifiziert werden können (lernfähige Algorithmen). Die symbolische Repräsentation des Verhaltens schließt die Lernfähigkeit also nicht aus, nur umgekehrt gibt es keine implizite Programmierung.
Ein wichtiges Beispiel sind Support Vector Machines (SVM), bei denen Datenmengen in höherdimensionalen Räumen abgebildet werden, um dort eine Gruppierung vornehmen zu können.[96] Das Resultat entspricht einem gelernten Modell der Umwelt. SVM sind wichtige Werkzeuge bei der Mustererkennung und werden z.B. bereits im medizinischen Bereich (zur Erkennung krankhaft veränderten Gewebes) eingesetzt.[97]

2. Neuronale Netze

Künstliche neuronale Netze (artificial neural networks) sind bereits Gegenstand der juristischen Literatur.[98] In ihnen wird Information – ähnlich wie bei natürlichen neuronalen Netzen – implizit durch den Zustand des gesamten Netzes repräsentiert. In einem solchen Netz sind die einzelnen künstlichen Neuronen jeweils mit zahlreichen (bis zu tausend) anderen verknüpft. Das einzelne Neuron wird durch eine besondere Funktion simuliert.[99] Der Input an das einzelne Neuron wird dabei unterschiedlich gewichtet

96 Dazu *Schölkopf/Smola*, Learning with Kernels, 2002, S. 23 ff.; *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 147 f.; *Kubat*, An Introduction to Machine Learning, 2017, S. 97 f.; *Zech*, ZfPW 2019, 198, 201.

97 *Klaus-Robert Müller*, persönliche Mitteilung, ; Sandkühler/Overhoff, Machbarkeit der Gebedifferenzierung in 3D-Ultraschallvolumina, 2014, https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0034-1389535 (zuletzt aufgerufen am 10.10.2011); Yang/Zhou/Yi/Chen/Chen, Journal of Medical Systems, Sep2019, Vol. 43 Issue 9.

98 *Ehinger/Stiemerling*, CR 2018, 761 ff.; *Söbbing*, K&R 2019, 164 ff.

99 Eine durch eine Funktion repräsentierte Schaltstelle ist also mit bis zu tausend anderen verknüpft. Die Komplexität eines neuronalen Netzes ergibt sich aus der Zahl der Schaltstellen und der Zahl

(weights) im Neuron aufsummiert und führt bei diesem Neuron, sobald ein bestimmter Grenzwert überschritten ist, zur Aktivierung, die dann wiederum einen Input an anderen Neuronen bewirkt. Sowohl die Gewichtungen als auch der Schwellenwert lassen sich verändern, wobei Algorithmen, die eine Veränderung aufgrund von Rückmeldevorgängen erlauben (Auslösefunktion, insbesondere backpropagation algorithm[100]) von besonderer Bedeutung sind. Die einzelnen Neuronen können in verschiedenen Schichten angeordnet werden, wobei ein „input layer", mehrere sogenannte „hidden layers" (verborgene Zwischenschichten) und ein „output layer" unterschieden werden können.[101] Die Gesamtheit der Gewichtungen und Schwellenwerte in einem bestimmten Netz repräsentieren implizit die Regeln, nach denen die Informationsverarbeitung ausgeführt wird.

Künstliche neuronale Netze wurden erst dadurch praktisch relevant, dass eine bestimmte Zahl von Neuronen technisch realisiert werden konnte.[102] Hinzu kam die mehrschichtige Architektur und Weiterentwicklungen der verwendeten mathematischen Elemente. Als dritter Faktor konnten neuronale Netze, wie lernfähige digitale Systeme überhaupt, erst dadurch erfolgreich werden, dass neben der großen Rechenleistung eine große Menge an Trainingsdaten verfügbar wurde (durch die Fortschritte in Sensorik, Vernetzung und Datenspeicherung – Big Data).

Ein Problem neuronaler Netze besteht darin, dass sie Information und damit auch das gelernte Verhalten nur implizit repräsentieren (das Ganze ist mehr als die Summe seiner Teile). Man spricht von der „Black Box".[103] Dieser Intransparenz neuronaler Netze versucht explainable AI (XAI) gegenzusteuern. Dabei geht es letzten Endes darum, symbolische Erklärungen für erlerntes Verhalten zu generieren. So kann etwa anhand einer Heatmap gezeigt werden, welche Bereiche von Bildern, die als Trainingsdaten verwendet wurden, mehr oder weniger zum Gelernten beigetragen haben. Auch lässt sich der Zustand einzelner Zwischenschichten abstrahieren und entsprechend darstellen (etwa als graphische Elemente mittlerer Abstraktheit bei der Bilderkennung).

---

der Verknüpfungen. Zu künstlichen neuronalen Netzen *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 157 ff.; *Kubat*, An Introduction to Machine Learning, 2017, S, 91 ff. Eine gute allgemeine Einführung findet sich bei *Lewis-Kraus*, The Great A.I. Awakening, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html (zuletzt aufgerufen am 10.10.2019).

100 Zur Rolle und Entwicklung des backpropagation algorithm *Flasiński*, Introduction to Artificial Intelligence, 2016, S. 169; *Kubat*, An Introduction to Machine Learning, 2017, S. 97 ff.

101 Von dieser mehrschichtigen Architektur rührt auch die Bezeichnung *deep learning*.

102 Die als Elemente von ANN erforderlichen Algorithmen, insbesondere den backpropagation algorithm, wurden bereits früher entwickelt. Die technische Umsetzung in hinreichend großen künstlichen Netzen gelingt erst, seit genügend Rechenleistung zur Verfügung steht und begann vor etwa zehn Jahren.

103 Dazu *Yuan*, RW 2018, 477, 483 f.; *Linardatos*, ZIP 2019, 504, 505; *Zech*, ZfPW 2019, 198, 202. S.o. Fn. 84.

3. Hybride Systeme

Wie bereits unter II.3 erwähnt sind hybride Systeme von hoher praktischer Relevanz. Gerade durch die Verbindung unterschiedlicher digitaler Systeme wird die Lösung komplexerer Aufgaben ermöglicht.

VIII. Spezifische Risiken lernfähiger (autonomer) Systeme: Vorhersehbarkeit und Erklärbarkeit des Verhaltens

Für das Haftungsrecht sind die besonderen Risiken, die mit der Lernfähigkeit verbunden sind, von Interesse. Man kann von einem *Autonomierisiko* sprechen.[104] Eine ausführliche Darstellung der Chancen und Risiken autonomer Systeme findet sich bei *Schulz*.[105]

Ex ante ist die Vorhersehbarkeit des Verhaltens dadurch, dass es erst gelernt wird, eingeschränkt, was eine verringerte Kontrolle mit sich bringt. Dadurch entsteht das Risiko eines Schadenseintritts durch unvorhersehbare Umstände. Hinzu kommt eventuell bei komplexen neuronalen Netzen die Möglichkeit spontanen Verhaltens. Ex post bedeutet Lernfähigkeit, dass es schwieriger wird, zu erklären, wie ein bestimmtes Verhalten zustande kam, das letztendlich zu einem Schaden geführt hat.

1. Vorhersehbarkeit (ex ante): Einfluss des Gelernten (verringerte Risikobeherrschung durch den Programmierer)

Die eingeschränkte Vorhersehbarkeit des Verhaltens stellt das zentrale Problem lernfähiger Systeme dar.[106] *Schirmer* formuliert in einem aktuellen Beitrag treffend: „Am Anfang steht das Autonomierisiko. Weil autonome Systeme nicht länger vorgegebene

104 *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 175; *Teubner*, AcP 218 (2018), 155, 164; *Cornelius*, ZRP 2019, 8, 9; *Schirmer*, RW 2018, 453, 465; *ders*., JZ 2019, 711.

105 *Schulz*, Verantwortlichkeit bei autonom agierenden Systemen, 2015, S. 72 ff.

106 *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 228; *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 37 f.; *Borges*, NJW 2018, 977, 978; *Hacker*, RW 2018, 243, 257 (mit Hinweis auf entsprechende Eigenschaft menschlichen Handelns); *Schirmer*, JZ 2019, 711; *Spindler*, in Lohsse/Schulze/Staudenmayer (Hrsg.), Liability for Artificial Intelligence and the Internet of Things,

Muster abarbeiten, sondern mit einem Entscheidungsspielraum ausgestattet werden, lässt sich im Vorfeld nicht mehr mit Sicherheit sagen, welche Entscheidung das System im konkreten Einzelfall treffen wird."[107]

Die Einschränkung der Vorhersehbarkeit unterscheidet sich für die verschiedenen Akteure.[108] Zunächst ist vor allem für den Entwickler bzw. Programmierer lernfähiger Systeme die Vorhersehbarkeit eingeschränkt. Je nachdem, welche Vorgaben (s.o. IV.) er macht, kann er das spätere Verhalten mehr oder weniger gut vorhersehen. Ihm stehen Garantien und Restriktionen als technisches Mittel der Beherrschung zur Verfügung.[109] Jedenfalls aber hat er nicht mehr, wie bei der klassischen Programmierung, vollständige Kenntnis über die Regeln, die das Verhalten ausmachen.

Als zusätzlicher Akteur tritt der Trainer auf den Plan (auf Hersteller- oder auf Anwenderseite). Für ihn ist das spätere Verhalten ebenfalls mehr oder weniger gut vorhersehbar, je nachdem für welche Lernmethode das System designt ist und welche sonstigen Vorgaben (s.o. V.) er machen kann (aufgrund der Vorgaben des Programmierers) und macht.

Als weitere Akteure kommen Nutzer und Betroffene ins Spiel, die keinen Einfluss auf das Verhalten haben. Auch für sie ändert sich die Vorhersehbarkeit. Der Umgang mit programmierten digitalen Systemen gehört mittlerweile zur normalen Alltagserfahrung, der Umgang mit lernfähigen Systemen noch nicht.[110] Daher ist die Definition von *Günther* für Autonomie interessant „Wird im Folgenden von einem hohen Grad von Autonomie gesprochen, soll dies zunächst bedeuten, dass der Betroffene wenig Wissen über das System hat, er also in der Entscheidungsvorhersage mit hoher Wahrscheinlichkeit nicht richtig liegen wird. Diese Unvorhersehbarkeit ist gerade im Rahmen der rechtlichen Verantwortung zu bewerten."[111] Allerdings gilt dies in erster Linie prospektiv. Retrospektiv lassen sich Geschehensabläufe auch mit der Hilfe von Fachleuten (Gutachter etc.) aufklären, so dass hier ein anderer bzw. allgemeinerer objektiver Maßstab anzulegen ist.

2019, S. 125, 126; *Riehm/Meier*, in DGRI Jahrbuch 2018 (erscheint demnächst), S. 3 f. Zum amerikanischen Recht *Calo*, 103 Cal. L. Rev. 513 (2015), 542.

107 *Schirmer*, JZ 2019, 711.

108 *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 37: „So erkennen alle Ansichten an, dass sich technische Systeme für *bestimmte* Betrachter unvorhersehbar verhalten können. Damit ist nicht gemeint, dass diese Systeme sich für *alle* Betrachter unvorhersehbar verhalten müssen."

109 *Reichwald/Pfisterer*, CR 2016, 208, 212. Malte Helmert

110 Hierher gehört auch die Diskussion um eine mögliche Anthropomorphisierung lernfähiger Systeme.

111 *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 38.

Grundsätzlich bedeutet Lernfähigkeit also immer verringerte Beherrschung durch denjenigen, der keinen Einfluss auf den Lernvorgang hat. Die Frage der Vorhersehbarkeit ist für jeden, der am Einsatz des lernfähigen Systems beteiligt ist (Entwickler, Trainer, Betreiber, Benutzer), unterschiedlich zu beurteilen. Dabei ist nicht nur nach dem Grad der Autonomie zu differenzieren, sondern insbesondere auch danach, in wessen Sphäre trainiert wird: Was wurde programmiert? Wer hat trainiert? Wie lange bzw. in welcher Phase (System auch noch auf dem Markt lernfähig oder nicht)?

2. Vorhersehbarkeit: Selbstorganisation und spontanes Verhalten

Das Verhalten lernfähiger Systeme wird gelegentlich auch als nicht deterministisch bezeichnet.[112] Ein deterministischer Algorithmus verhält sich vollkommen vorhersagbar, auch hinsichtlich der Zwischenschritte.[113] Demgegenüber arbeiten probabilistische Algorithmen mit Zufallszahlen, sind also hinsichtlich ihrer Zwischenschritte nicht vollkommen vorhersagbar, um aber ein zumindest mit einer gewissen Wahrscheinlichkeit vorhersagbares Ergebnis zu erzielen. Hinsichtlich des Ergebnisses sind also auch probabilistische Algorithmen determiniert. Lernfähigkeit kann dagegen je nach Trainingsdaten zu wirklich unvorhersehbaren Lernergebnissen führen.

Der Einsatz probabilistischer Verfahren ist aus rechtlicher Sicht kein Problem, da sich auch das Recht mit Wahrscheinlichkeiten zufriedengibt, wenn diese hinreichend sind (hinreichende Sicherheit). Nicht-deterministisches Verhalten fügt sich in die vorhandene rechtliche Risikodogmatik problemlos ein, da diese auf Abwägung von Chancen und Risiken basiert, mithin also eine statistische Betrachtung vornimmt.[114] Das gleiche gilt für die Technik (Technikfolgenabschätzung) und sogar für die Wissenschaft (erkenntnistheoretische Frage, wann eine Erkenntnis als gesichert gelten darf; Einfluss von Wahrscheinlichkeitsberechnungen auch in der Erkenntnistheorie, etwa im Bayesianismus).

Eine andere Frage ist, ob komplexe lernfähige Systeme darüber hinaus inhärent unvorhersehbares Verhalten zeigen können. Selbstorganisation kann zu Eigendynamik

112 *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 228; *Zech*, ZfPW 2019, 198, 202; *Riehm/Meier*, in DGRI Jahrbuch 2018 (erscheint demnächst), S. 3 f.

113 S.o. A.III.1. *Grützmacher*, CR 2016, 695, 697 spricht von „*Turing*-Berechenbarkeit", *Reichwald/Pfisterer*, CR 2016, 208, 211 von „algorithmischer Prädeterminierung".

114 *Mazzini*, A System of Governance for Artificial Intelligence through the Lens of Emerging Intersections between AI and EU Law, SSRN Nr. 3369266, S. 13: „the ‚unpredictability' of AI does not appear to represent a cause of new concern as a matter of principle from a safety standpoint".

und Spontaneität führen.[115] Damit könnten etwa komplexe neuronale Netze nach dem Vorbild natürlicher Gehirne spontanes Verhalten zeigen.[116] Solche komplexe lernfähige Systemen, die zur Selbstorganisation fähig sind und dadurch vergleichbar mit höheren Tieren Eigendynamik und Spontaneität aufweisen, sind aber im Bereich der Alltagsrobotik noch Zukunftsmusik.

Neuronale Netze verhalten sich grundsätzlich nicht besonders, auch sie sind zunächst nur aus Algorithmen aufgebaut und führen daher bei exakt gleichen Eingaben zu exakt gleichen Ausgaben. Möglicherweise kann es aber bei komplexen neuronalen Netzen mit Rückkoppelung[117] zu inhärent unvorhersehbaren Effekten kommen. Selbst ein fest programmiertes bzw. deterministisches System kann sich bei iterativer Anwendung unvorhergesehen verhalten (was Gegenstand der Chaostheorie ist).[118] Bereits sehr kleine Abweichungen beim Input können dann zu großen Abweichungen im Verhalten führen. Umgekehrt wird auch bei neuronalen Netzen erforscht, ob die Verwendung von Zufallseingaben (probabilistischer Algorithmus) das Verhalten des Systems insgesamt stabilisieren kann.[119]

Ähnliche Überlegungen könnten auch bei Schwärmen aus digitalen Systemen zutreffen.[120] Auch diese sind aber im praktischen Alltag noch Zukunftsmusik. Bislang jedenfalls wurde ein mögliches chaotisches Verhalten digitaler Systeme aus rechtlicher Sich noch nicht ausführlich diskutiert. Denkbar wäre, dass auch ein besonderes Risiko chaotischer digitaler Systeme zum Thema wird, aktuell ist aber auch dies noch ebenso Zukunftsmusik wie Überlegungen zu KI mit eigenem Bewusstsein bzw. Persönlichkeit.

---

115 *Klaus Mainzer*, Vortrag am 5.10.2015, Universität Passau.

116 *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 170.

117 Etwa bei adversarial neural networks, d.h. wenn man zwei neuronale Netzwerke miteinander bzw. gegeneinander trainieren lässt.

118 Ein einfaches Beispiel ist „Langtons Ameise“, vgl. *Stewart*, Spektrum der Wissenschaft 1995, 10. Für den Hinweis danke ich *Malte Helmert*.

119 *Manfred Hauswirth*, persönliche Mitteilung.

120 *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 227: Möglichkeit der „ad hoc-Vernetzung“ von Softwareagenten zu Multiagentensystemen. Vgl. *Kaplan*, Künstliche Intelligenz, 2015, S. 66. Zur Schwarmrobotik s.o. B.V.3.

3. Erklärbarkeit (ex post)

Der eingeschränkten Vorhersehbarkeit ex ante, entspricht die eingeschränkte Erklärbarkeit ex post.[121] Mit Hilfe der Explainable AI (s.o. III.2) lassen sich aber zumindest abstrahierte, mit einer gewissen Wahrscheinlichkeit versehene Aussagen über die Ursachen für ein bestimmtes Verhalten machen.

Auch hier ist zwischen der Intransparenz für durchschnittliche Anwender[122] (Verbraucher) und der Intransparenz auch für den Fachmann, zu unterscheiden.[123] Anders als bei der Betrachtung ex ante ist aber nur letztere ein Problem, da im Schadensfall bzw. bei dessen rechtlicher Aufarbeitung Gutachter eingesetzt werden können (vgl. Verständnis der Funktionsweise eines Autos oder eines Computers für den durchschnittlichen Verbraucher, das nur auf einer hohen Abstraktionsebene gegeben ist, und Einsatz entsprechender Gutachter).

Ein Problem der Nachweisbarkeit liegt darin, dass zur Erklärung des Gelernten auch die Trainingsdaten mitgeloggt bzw. protokolliert werden müssten.[124] Dies kann bei Sensordaten zu einem Kapazitätsproblem führen.[125] Möglicherweise lassen sich aber auch aggregierte Daten speichern, die aus mitlaufenden Diagnoseverfahren (Explainable AI) gewonnen werden.

E. Vernetzung: cyber-physikalische Systeme

Als dritter wichtiger Aspekt neben Robotik und Lernfähigkeit soll die zunehmende Vernetzung angesprochen werden.

---

121 Dagegen lässt sich das Verhalten von Algorithmen zumindest theoretisch ex post vollständig erklären. *Reichwald/Pfisterer*, CR 2016, 208, 211: „Aktuelle Systeme, die zwingend algorithmisch arbeiten, können noch keine echten ‚bewussten' Entscheidungen treffen und deren Konsequenzen abschätzen. Die Entscheidungen sind aber aufgrund der algorithmischen Prädeterminierung retrospektiv unter Kenntnis der Historie und des Algorithmus nachvollziehbar." Voraussetzung ist allerdings auch hier, dass die gesamte Historie bekannt ist, also der gesamte Input, der zu einer bestimmten Reaktion eines Systems geführt hat.

122 *Günther*, Roboter und rechtliche Verantwortung, 2016, S. 38, stellt auf den „Betroffenen" ab.

123 Vgl. *Martini*, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, 2019, S. 28, der auf die „Sicht des durchschnittlichen Nutzers einer Softwareanwendung" abstellt.

124 *Horner/Kaulartz*, CR 2016, 7, 10.

125 *Reichwald/Pfisterer*, CR 2016, 208, 211.

I. Vernetzung digitaler Systeme untereinander: vom Internet zum Internet of Things

Das seit den 1990ern bestehende Internet ist bekannt und klassisches Beispiel einer digitalen transformativen Technologie. Dabei geht es aber nur um den Informationsaustausch, also die Vernetzung von Menschen bzw. virtueller digitaler Systeme. Neu ist die Vernetzung von Sensoren und Aktuatoren bzw. von Robotik. Die virtuelle Vernetzung (Internet) wird auf physische Gegenstände ausgedehnt (Internet of Things, Internet der Dinge).

Die Vernetzung körperlicher Gegenstände steht aktuell erst am Anfang. Das Internet of Things[126] bzw. cyber-physikalische Systeme[127] finden aber zunehmend Verbreitung. *Möller* beschreibt das Internet of Things anschaulich als „networked interconnection of everyday objects".[128] Die Vernetzung körperlicher Gegenstände soll also nicht nur als Industrie 4.0 im unternehmerischen Bereich stattfinden, sondern im gesamten Alltag Einzug halten. Damit wird bei den meisten komplexen digitalen Analyse- und Steuerungssystemen auch der Aspekt der Vernetzung zum Tragen kommen.

II. Vernetzung von Sensoren und Aktuatoren, Funktechnik

Das Internet of Things betrifft vor allem die Vernetzung von Sensoren, weshalb es auch bereits im Kontext von Big Data eine große Rolle spielte. Bereits bei einfachen Haushaltsgeräten kommen aber bereits Aktuatoren hinzu, dies gilt erst recht in der Industrie 4.0 und bei vernetzten Robotern. Die Vernetzung betrifft also sowohl die Eingabeseite (Vernetzung von Sensoren) als auch die Ausgabeseite (Vernetzung von Robotern).

---

126 Das Internet of Things (IoT) kann definiert werden als „global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies" (International Telecommunication Union, Recommendation ITU-T Y.2060, 3.2.3). Zum Begriff *Hornung/Hofmann*, in Hornung (Hrsg.), Rechtsfragen der Industrie 4.0, 2018, S. 9 f. Vgl. auch *Sosnitza*, CR 2016, 764, 765.

127 Eine Definition cyber-physikalischer Systeme im Zusammenhang mit der rechtlichen Diskussion findet sich bei *Reichwald/Pfisterer*, CR 2016, 208: „Sie sind als Systeme definiert, die physische Komponenten mit Software verflechten, wobei jedes CPS in unterschiedlichen zeit- und räumlichen Bereichen operieren kann, unterschiedliche Modalitäten und Verhaltensweisen aufweist und in vielfältiger Weise mit anderen Systemen kontextabhängig kommunizieren und interagieren kann." Vgl. *Möller*, Guide to Computing Fundamentals in Cyber-Physical Systems, 2016, S. 81 ff.

128 *Möller*, Guide to Computing Fundamentals in Cyber-Physical Systems, 2016, S. 141.

Für die Vernetzung frei beweglicher Hardwaresysteme ist die Funktechnologie von großer Bedeutung. Wie bei den Sensoren und Aktuatoren trägt hier die Miniaturisierung und Steigerung der Energieeffizienz zur Verbreitung bei (vgl. B.III zur Robotik, Beispiele für effiziente Funktechnologien sind Bluetooth und RFID). Ein Technologiesprung dürfte hier durch die Einführung des 5G-Standards zu erwarten sein, der auch die Datenübertragungsrate (Datenvolumen pro Zeit) erhöht. Durch die Fortschritte in der Funktechnik wird auch die Vernetzung frei beweglicher Roboter möglich (insbesondere auch selbstfahrender Fahrzeuge[129]).

III. Vernetzung und Bestimmbarkeit einzelner Systeme

Eine weitere Folge der zunehmenden Vernetzung ist, dass die Bestimmbarkeit bzw. Abgrenzbarkeit einzelner digitaler Systeme erschwert wird. Damit wird auch die Bestimmung einzelner Akteure, die für ein bestimmtes System verantwortlich sind, schwieriger.

Allerdings bleibt bei physischen Systemen immer die Bestimmung der Hardware möglich. Hat ein Roboter einen Schaden verursacht, so lässt sich jedenfalls dieser Roboter genau bestimmen (schwieriger wird es bei Infrastrukturrobotik). Zwar lassen sich wegen der Vernetzung möglicherweise die Ursachen für das Verhalten des Roboters nicht bestimmen, der Roboter selbst als Schadensquelle steht aber als Anknüpfungspunkt für rechtliche Regelungen zur Verfügung.

IV. Spezifische Risiken der Vernetzung

Die Vernetzung dürfte das Haftungsrecht vor die größten Herausforderungen stellen. Erst durch die Vernetzung wird die Bestimmung einzelner Verursachungsbeiträge so schwierig, dass das auf einer kausalen Schadensverursachung durch einzelne Akteure aufbauende Haftungsrecht an seine Grenzen kommt. Fremde Daten können unabsichtlich oder absichtlich zu Schädigungen führen. Man kann daher von einem eigenen *Vernetzungsrisiko* sprechen.[130]

129 Zu den Szenarien der Datenströme im Straßenverkehr mit und ohne Plattformen *Kerber/Frank*, Data Governance Regimes in the Digital Economy: The Example of Connected Cars, SSRN Nr. 3064794.

130 *Zech*, in Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 163, 175; *ders*.

1. Schädigung durch fremde Daten (ungewollt)

Das Vernetzungsrisiko ergibt sich insbesondere aus der Möglichkeit der Schädigung durch ungewollt fehlerhafte Daten aus dem Netz (safety risk) und aus der Ermöglichung von Angriffen über das Netz (security risk, dazu unter 2.). Das Risiko ungewollter Schäden entsteht wegen hinzukommender Fernwirkungsszenarien. Durch die Vernetzung können Probleme in einer Vielzahl von Datenquellen zu Problemen bei einem System führen. Handelt es sich nicht um virtuelle Systeme, kann dies auch unmittelbar zu einer Schadensentstehung führen.
Für den Nachweis ex post gilt das bereits unter C.VII.3 Ausgeführte. Das Mitloggen bzw. Protokollieren kann zu Kapazitätsproblemen führen, die aber möglicherweise technisch zu bewältigen sind.

2. Angreifbarkeit

Besonders problematisch ist, dass eine stärkere Vernetzung auch die Angreifbarkeit erhöht. Es entstehen zusätzliche Angriffs- bzw. Missbrauchsmöglichkeiten für vorsätzlich handelnde Dritte.[131] Dieses Risiko missbräuchlicher Eingriffe, kombiniert mit der Möglichkeit einer unmittelbaren physischen Schädigung, hält *Bruce Schneier* für die größte aktuelle Herausforderung im Umgang mit digitalen Systemen.[132]
Technische Möglichkeiten der Absicherung („security by design“) finden sich auch auf der Softwareseite (etwa Kryptographie). Es handelt sich um ein typisches Problem der IT-Sicherheit mit den drei Hauptzielen Vertraulichkeit, Verfügbarkeit und Integrität. Dass dies mangelnde IT-Sicherheit auch unmittelbare physische Schäden zur Folge haben kann, wurde rechtlich bislang nur für kritische Infrastrukturen[133] berücksichtigt.

ZfPW 2019, 198, 205; *Teubner*, AcP 218 (2018), 155, 164, 201 ff.; *Cornelius*, ZRP 2019, 8, 10; *Riehm/Meier*, in DGRI Jahrbuch 2018 (erscheint demnächst), S. 13. Zu den haftungsrechtlichen Konsequenzen *Spiecker gen. Döhmann*, CR 2016, 689 ff.; *Wende*, in Sassenberg/Faber (Hrsg.), Rechtshandbuch Industrie 4.0 und Internet of Things, 2017, S. 69; *Hornung/Hofmann*, in Hornung (Hrsg.), Rechtsfragen der Industrie 4.0, 2018, S. 9, 29 ff.; *Spindler*, in Lohsse/Schulze/Staudenmayer (Hrsg.), Liability for Artificial Intelligence and the Internet of Things, 2019, S. 125, 127 f.

131 *Hilgendorf*, RAW 2018, 85, 88. Dabei handelt es sich um ein Problem der sog. Cybersicherheit, vgl. speziell zu Industrie 4.0 *Möller*, Guide to Computing Fundamentals in Cyber-Physical Systems, 2016, S. 336 ff.

132 *Schneier*, Click Here to Kill Everybody, 2018, 3.

133 §§ 8a ff. BSI-Gesetz; Richtlinie 2008/114/EG des Rates vom 8. Dezember 2008 über die Ermittlung und Ausweisung europäischer kritischer Infrastrukturen und die Bewertung der Notwendigkeit, ihren Schutz zu verbessern. Vgl. *Spiecker gen. Döhmann*, CR 2016, 689, 700.

3. Komplexe Verursachung

Neben der Höhe des Risikos ist aus haftungsrechtlicher Sicht wichtig, dass durch Vernetzung (und Lernfähigkeit) Szenarien drohen, bei denen an einer Schadensentstehung sehr viele Akteure beteiligt sind. Durch die Vernetzung können beliebig viele Personen hinzukommen, „deren" Systeme Daten geliefert haben (wobei sich hier wieder dieselben Zurechnungsfragen stellen). *Calo* spricht anschaulich von einer Datenpromiskuität („data promiscuity"),[134] *Balkin* von der „work of many hands"[135].
Hinzu kommt, dass vernetzte Systeme zu einem komplexen Gesamtsystem werden können, das unvorhersehbares Verhalten zeigt (s.o. C.VII.2).[136] Hardwaresysteme können die Fähigkeit erlangen, als Schwarm zu agieren, wodurch ebenfalls ein unvorhersehbares emergentes Verhalten möglich wird.[137]

F. Auswirkungen

Nur kurz sollen mögliche Auswirkungen dargestellt werden. Drei zentrale Themen sind die Entstehung neuer Risiken, die Verlagerung von Einflussmöglichkeiten und der Vergleich digitaler Systeme mit Menschen.
Ein klassisches Thema von Technikphilosophie und Technikrecht ist die Entstehung neuer Risiken durch neue Technologien. Nichts anderes ist auch für Robotik, Lernfähigkeit und Vernetzung zu konstatieren. Sie schaffen gegenüber herkömmlichen digitalen Systemen neue Möglichkeiten des Schadenseintritts und damit neue Risiken. Wie bei anderen neuartigen Technologien auch, ist die Höhe der Risiken, also ihre Eintrittswahrscheinlichkeit und die Höhe der möglichen Schäden, nur ungenau abzuschätzen. Da es sich aber unbestreitbar um eine äußerst nützliche Technologie handelt, muss für einen angemessenen Umgang mit den Risiken gesorgt werden.
Mit den Risiken verbunden ist die Frage, wer Einfluss auf sie hat. Dabei treten durch Lernfähigkeit und Vernetzung neue Akteure auf den Plan, die bekannte ergänzen. So tritt neben den Programmierer der Trainer, der ebenfalls Einfluss auf das Verhalten eines lernfähigen digitalen Systems hat. Bei vernetzten Systemen kann durch Daten-

134 *Calo*, 103 Cal. L. Rev. 513 (2015), 555: „data promiscuity"; vgl. *Balkin*, 6 Cal. L. Rev. Cir. 45 (2015), 54.

135 *Balkin*, 6 Cal. L. Rev. Cir. 45 (2015), 53.

136 *Kirn/Müller-Hengstenberg*, MMR 2014, 225, 228; *Hornung/Hofmann*, in Hornung (Hrsg.), Rechtsfragen der Industrie 4.0, 2018, S. 9, 31.

137 *Kaplan*, Künstliche Intelligenz, 2015, S. 66.

übermittlung eine unübersehbare Zahl anderer Akteure Einfluss auf das System nehmen. Dies mag zwar nicht automatisch zu einer Verantwortungsdiffusion führen, zumindest aber werden entferntere Verursachungen schwerer aufzuklären sein.

Besonders kontrovers diskutiert wird die Frage, ob und in welcher Hinsicht lernfähige digitale Systeme mit Menschen verglichen werden können. Aus der Sicht des Rechts bedeutet dies nicht zwingend eine Diskussion um die volle Anerkennung als Person. Vielmehr kennt das Recht auch Entitäten, denen nur für bestimmte Aspekte Rechtspersönlichkeit zugebilligt wird, etwa als Haftungssubjekt. Ebenfalls wird diskutiert, ob unabhängig von einer eigenständigen Anereknnung als Rechtsperson zumindest solche Regelungen Anwendung finden sollen, die das Handeln bzw. das Verhalten von Menschen (Hilfspersonen) anderen zurechnen. Solche Zurechnungsregeln finden sich bei Haftungsfragen (Zurechnung schädigenden Verhaltens) ebenso wie bei Erklärungen (Zurechnung von Willenserklärungen durch Stellvertretung) oder bei der Schöpfung von Immaterialgütern (so die „work made for hire“-doctrine im US-amerikanischen Urheberrecht). Allerdings stellt sich dabei immer die Frage, ob nicht die Behandlung digitaler Systeme als Werkzeug statt als Hilfsperson ebenfalls und auf einfacherem Weg zu zweckmäßigen Ergebnissen führt. Geht es um die Zurechnung der Schadensverursachung, steht die Gefährdungshaftung zur Verfügung, geht es um die Zurechnung der Schaffung von Immaterialgütern der Gedanke der Leistungsschutzrechte.