

Hobe u.a. (Hrsg.)

Die Macht der Algorithmen



Nomos

Schriften zu Recht und Ethik
der Digitalen Transformation

Herausgegeben von

Prof. Dr. Thomas Grundmann

Prof. Dr. Johanna Hey

Prof. Dr. Dr. h.c. Dr. h.c. Stephan Hobe

Prof. Dr. Christian Katzenmeier

Prof. Dr. Torsten Körber

Dr. Claes Neuefeind

Prof. Dr. Markus Ogorek

Prof. Dr. Dr. Frauke Rostalski

Prof. Dr. Dr. h. c. Martin Waßmer

Band 1

Thomas Grundmann | Johanna Hey | Stephan Hobe
Christian Katzenmeier | Torsten Körber | Claes Neuefeind
Markus Ogorek | Frauke Rostalski | Martin Paul Waßmer (Hrsg.)

Die Macht der Algorithmen



Nomos

Die Open-Access-Veröffentlichung dieses Titels wurde durch die Dachinitiative „Hochschule.digital Niedersachsen“ des Landes Niedersachsen ermöglicht.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2023

© Die Autor:innen

Publiziert von
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-7560-0622-9

ISBN (ePDF): 978-3-7489-4157-6

DOI: <https://doi.org/10.5771/9783748941576>



Onlineversion
Nomos eLibrary



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Vorwort

Ringvorlesung Die Macht der Algorithmen

Dieser Band versammelt sieben Vorträge, die im Wintersemester 2021/22 und im Sommersemester 2022 im Rahmen einer von der Forschungsstelle „Recht und Ethik der digitalen Transformation“ veranstalteten Ringvorlesung unter dem Rahmenthema „Die Macht der Algorithmen“ gehalten wurden.

Er ist damit zugleich Auftaktband der Schriftenreihe der seit 2020 bestehenden Forschungsstelle, in der fächerübergreifend Wissenschaftlerinnen und Wissenschaftler der Rechtswissenschaftlichen, der Wirtschafts- und Sozialwissenschaftlichen sowie der Philosophischen Fakultät der Universität zu Köln aktiv sind.

Wir sind den Dekaninnen und Dekanen der drei Fakultäten dankbar für das Verständnis, welches sie dem Vorhaben entgegengebracht haben.

Herrn Wiss.Mit. Sebastian Tober danken wir für die technische Unterstützung bei der – zumeist noch gemäß Corona-Bedingungen – abgehaltenen Ringvorlesung. Dem Nomos Verlag und hier insbesondere Herrn Dr. Marco Ganzhorn ist für die professionelle Umsetzung im Rahmen dieser neuen Schriftenreihe zu danken.

Köln, im Januar 2023

Thomas Grundmann, Johanna Hey, Stephan Hobe, Christian Katzenmeier, Torsten Körber, Claes Neufeind, Markus Ogorek, Frauke Rostalski und Martin Paul Waßmer

Inhaltsverzeichnis

<i>Nina Eckertz, Øyvind Eide</i> Don't fear Black-Clouds – Mechanismen künstlicher Intelligenz	9
<i>Stephan Hobe, Martin Schwaborn</i> Völkerrechtliche Schwierigkeiten einer effektiven Bekämpfung von Cyber-Angriffen	31
<i>Amina Hoppe</i> Der Einfluss von Social Bots auf Debatten in sozialen Netzwerken am Beispiel Twitter	61
<i>Christian Katzenmeier</i> Haftung für Schäden durch KI in der Medizin	73
<i>Torsten Körber</i> Google, Facebook & Co: Die Macht von Daten und Algorithmen im Fokus des Kartellrechts	103
<i>Axel Ockenfels</i> ... denn sie wissen, was wir tun	121
<i>Martin Paul Waßmer</i> Auf dem Sprung zum autonomen Fahren	127
Kurzbiografien der Autorinnen und Autoren	149

Don't fear Black-Clouds – Mechanismen künstlicher Intelligenz

Nina Eckertz, Øyvind Eide

A. Einleitung

Künstliche Intelligenz (KI) begegnet uns an zahlreichen Stellen im Alltag. Egal ob als Empfehlungsalgorithmus auf einer Social-Media-Plattform, in Form eines Saugroboters oder der Berechnung des schnellsten Weges zur Arbeit. Obwohl zumindest Teilaspekte der KI bereits vollständig in unserer alltäglichen Routine angekommen sind, ist nicht jeder positiv auf KI zu sprechen,¹ denn häufig ist es nicht direkt ersichtlich, wo, wie und welche Art von Daten verarbeitet werden oder wie genau die Methodik des KI-Systems funktioniert.

Dieses Phänomen der Nicht-Nachvollziehbarkeit über die Funktionalität des KI-Mechanismus kann mithilfe des Black-Box-Begriffs beschrieben werden. In der Systemtheorie bezeichnet „Black-Box“ ein nur durch das äußere Verhalten bekanntes System. So wird eine Black-Box allgemein als ein technischer Begriff ohne starke negative oder positive Konnotationen verstanden. Im Kontext des KI-Diskurses wird eine Black-Box jedoch oft als ein Begriff mit negativen Konnotationen empfunden. Die Kernfrage hinter der Black-Box im Rahmen von künstlicher Intelligenz lautet: Wie trifft eine künstliche Intelligenz eine Entscheidung? Um sich diesem Phänomen anzunähern, beschäftigt sich dieser Beitrag mit dem Black-Box-Begriff und der Nachvollziehbarkeit von KI in verschiedenen Paradigmen. Ziel dieses Beitrags ist es, einen Teil zur Aufklärung über KI beizutragen und zur Diskussion anzuregen. Dazu werden die Hintergründe des Black-Box-Begriffs sowohl mithilfe eines historischen Beispiels als auch den biologischen Vorbildern und deren informationstechnischer Umsetzung betrachtet. Hierzu werden zusätzlich aktuelle Modellierungen innerhalb der KI-Forschung geisteswissenschaftlich hinterfragt.

Hieraus ergibt sich, dass es viele unterschiedliche Faktoren gibt, die das Verständnis von KI-Systemen beeinflussen und deshalb der statische

1 TÜV-Verband e.V. (Hrsg.), Sicherheit und Künstliche Intelligenz, Erwartungen, Hoffnungen, Risiken. August 2021, S. 5.

Begriff der Black-Box unzureichend ist, um das Nicht-Nachvollziehbarkeit-Phänomen zu beschreiben. Diese unterschiedlichen Faktoren umfassen das persönliche Verständnis von KI-Systemen im alltäglichen Umgang, das gezielte *overselling* und *overbuying* von KI-Systemen und die technische Komplexität von KI-Systemen, angefangen bei der Modellierung von künstlicher Intelligenz bis hin zur praktischen Umsetzung.

Letztendlich kommen wir zu dem Schluss, dass jeder ein individuelles Verständnis von KI hat und sich KI anders erklärt. Aus diesem Grund möchten wir mit dem Black-Cloud-Begriff einen flexibleren Terminus für die Erfassung des Verständnisses von KI-Systemen vorschlagen. Die Idee hinter der Black-Cloud ist dabei, dass das Verständnis von KI sich nicht fest in einer Box messen lässt, sondern sich wie eine Wolke über die Zeit hinweg verändert. Dieser neue Begriff soll daher sowohl die individuelle Perspektive als auch zukünftige Entwicklungen abdecken und mehr Flexibilität bieten.

B. Black-Box-Begriff – Eine historisch basierte Einführung

Dieser Beitrag bezieht sich mit dem Begriff der „Künstlichen Intelligenz“ grundlegend auf eine Forschungsdisziplin innerhalb der Informatik, in welcher man die (menschliche) Intelligenz mithilfe verschiedener Technologien, z.B. komplexer Datenverarbeitung oder der Robotik, zu imitieren versucht.

KI bietet eine Vielzahl an Unterkategorien, wie bspw. *maschinelles Lernen* (engl. machine learning) sowie *computer vision* oder *natural language processing*, um die KI-Anwendung in den Kontext eines bestimmten Input-Mediums zu setzen. Diese Technologien werden dabei gerne im Kontext der Data Science betrachtet und benutzt. Je nachdem wie vielfältig die Fähigkeiten eines KI-Systems sind, kann zwischen einer schwachen und einer starken KI unterschieden werden.

Gemäß Russell und Norvig ist KI historisch betrachtet ein recht junges Feld innerhalb der Wissenschaft, denn die intensive Beschäftigung mit der KI startete erst kurz nach dem 2. Weltkrieg.² Heute, rund 70 Jahre später, ist es noch nicht gelungen, (menschliche) Intelligenz vollständig zu modellieren und technisch nachzubilden. Es ist also noch viel Platz für weitere „Edisons“ und „Einsteins“ innerhalb der KI-Forschung.³ Im Folgenden geben wir einen

2 S. Russell/P. Norvig, *Artificial intelligence: a modern approach*. Third, global edition, Boston: Pearson 2016, S. 16.

3 Russell/Norvig, *Artificial intelligence* (Fn. 2), S. 1.

Einblick in den aktuellen Diskurs der KI-Forschung und erklären das Phänomen Black-Box etwas detaillierter. Hierzu wird zunächst die Wahrnehmung von KI beleuchtet.

I. Wahrnehmung von KI – Black-Box-Hintergründe

Wie eingangs erwähnt, beziehen wir uns mit dem Begriff der künstlichen Intelligenz zunächst auf die Definition aus der Informatik. Im öffentlichen Diskurs ist KI jedoch weit mehr als eine reine Forschungsdisziplin. Anna Visvizi betont zum Beispiel, dass gerade im öffentlichen Diskurs KI sowohl als unendliche Chance als auch als Quelle für Risiko betrachtet werden würde.⁴ Man diskutiert hier oftmals nicht nur den aktuellen Forschungsstand und aktuelle Produkte, sondern auch die kurz- und langfristigen Folgen der KI-Forschung. Das liegt daran, dass in dieser Diskussion viele unterschiedliche Parteien teilnehmen, seien es Forschende, Politiker oder Wirtschaftsvertreter. Zusätzlich sind viele Bereiche betroffen, zum Beispiel die öffentliche Verwaltung, die Justiz oder das Militär. Ein wichtiger Aspekt rund um den KI-Diskurs ist daher nicht nur, in welchem thematischen Kontext KI diskutiert wird, sondern auch, wer an der Diskussion teilnimmt, wie viel (Vor-)Wissen man individuell mitbringt und wie voreingenommen, also *biased*, man ist.

Fortuna und Gorbaniuk schlagen daher beispielsweise in ihrer Studie zur Untersuchung, was Diskursteilnehmer mit dem „buzzword“ KI meinen, zwei Unterscheidungsmöglichkeiten der Sichtbarkeit von KI und von Nutzergruppen vor.⁵ Dies kann helfen, die Diskutanten erst mal grundsätzlich einzuordnen. Zum einen gibt es bei Fortuna und Gorbaniuk die unsichtbare KI, z.B. Verarbeitungsalgorithmen auf Social-Media-Plattformen, und die sichtbare KI, z.B. Sprachassistenten.⁶ Zum anderen bieten sie die Unterscheidung in IT-Experten und Laien an. Die Unterscheidung ergibt sich hierbei

4 A. Visvizi/M. Bodziany (Hrsg.), *Artificial Intelligence and Its Contexts: Security, Business and Governance. Advanced Sciences and Technologies for Security Applications*. Cham: Springer International Publishing, 2021, S. 13 f.; <https://link.springer.com/10.1007/978-3-030-88972-2> (besucht am 30.09.2022).

5 P. Fortuna/O. Gorbaniuk. What Is Behind the Buzzword for Experts and Laymen: Representation of „Artificial Intelligence“ in the IT-Professionals' and Non-Professionals' Minds, in: *Europe's Journal of Psychology* 18 (2022), 207 (209).

6 Fortuna/Gorbaniuk, *Buzzword* (Fn. 5), 207 (208 f.).

aus der beruflichen Qualifikation.⁷ Hierzu wurde auch manuell überprüft, wie viel Wissen die Studienteilnehmer über KI haben, mit dem Ergebnis, dass die IT-Experten deutlich mehr über KI wissen.⁸ Die Forscher kommen zu dem Schluss, dass es durchaus Unterschiede gibt, was mit KI gemeint ist. Zum einen ist mit KI bei Laien eher eine Imitation gemeint, zum anderen treffen die IT-Experten mehr Unterscheidungen von KI, z.B. nach Art von KI-Systemen und daher nach unterschiedlichen Anwendungsfällen.⁹ Das erweckt den Eindruck, dass es unterschiedliche Auffassungen des KI-Begriffs gibt.

Die Limitation dieser Studie liegt in ihrer Vereinfachung des bzw. der sichtbaren/unsichtbaren KI, aber auch in der Einordnung der Studienteilnehmer.¹⁰ Denn KI aus Perspektive der IT-Branche ist wie jedes Software-Projekt eine Teamleistung. Somit bietet die IT-Branche innerhalb der KI-Entwicklung nicht nur viele unterschiedliche Rollen, sondern auch vielfältige Aufgaben innerhalb dieser an. Eine feinkörnige Unterscheidung ist daher notwendig, um z.B. technische Rollen von semi- oder gar nicht-technischen Rollen abzugrenzen, denn diese bieten durchaus andere Perspektiven und anderes Vorwissen in Bezug auf den KI-Diskurs an. So macht es einen Unterschied, ob man sich mit einem Product Owner oder einem reinen Softwareentwickler austauscht, abhängig von deren Hintergründen.

Schaut man zunächst in die popkulturelle Darstellung und deren Narrative von künstlicher Intelligenz wie etwa der Androidin Ava in *Ex Machina*, dem *Terminator* oder Arisa in *Better Than Us*, zeigen viele Science Fiction-basierte Filme uns ein Bild auf, welches aus Sicht der heutigen Datenverarbeitung noch weit in der Zukunft oder eventuell gar nicht zu erreichen ist. So zeigen die genannten Exempel bspw. nur in sehr begrenzten Umfang den Schöpfungsprozess der KI auf und erklären nicht die technische Zusammensetzung ihrer KI.

Ein ähnlicher Ansatz ist in der Vermarktung von KI-Produkten zu beobachten. KI als Marketingbegriff und Buzzword ist bereits seit einigen Jahren in aller Munde, man kann hier durchaus von einem „Overselling“ sprechen. Ähnlich wie bei den Beispielen aus der Popkultur wird hier der KI-Begriff im Hype als „nie“ neue Technologie besprochen, wohingegen bspw. die Subkategorien und eigentlichen Innovationstreiber wie maschine-

7 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (209).

8 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (211).

9 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

10 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

lles Lernen nicht besprochen werden.¹¹ Das führt dazu, dass eine Lücke zwischen dem öffentlichen Bild von KI und den eigentlichen Fähigkeiten aktueller KI entsteht, wodurch Unternehmen bspw. ihren Einsatz von KI überdenken und ggf. davon absehen.¹²

Hierbei ist zu beachten, dass, aus wirtschaftlicher Sicht betrachtet, die zugrundeliegenden Technologien aufgrund des Betriebsgeheimnisses nicht genauer erklärt werden müssen. Auf rechtlicher Seite ist es gemäß Bibal et al. so, dass KI nur im öffentlichen Sektor eine Erklärung ihrer Entscheidungen bedarf, im privaten Sektor, egal ob Business-to-Business oder Business-to-Customer nur dann, wenn es ein spezifisches Gesetz dazu gibt.¹³ Das ist nicht nur im Sinne der Einordnung von „sichtbaren“ KI-Produkten (nach Definition von Fortuna/Gorbaniuk¹⁴) relevant, sondern gerade von unsichtbaren Produkten zum Personalisieren eines Produktes wie Empfehlungsalgorithmen auf sozialen Netzwerken. Insofern ist es nicht verwunderlich, wenn auf den ersten Blick intransparent ist, wie sich die KI hier zusammensetzt und wie sie ihre Entscheidungen trifft. Zusätzlich gibt es unterschiedliche Abstufungen für Transparenz und Erklärungsbedarf je nach Algorithmik-System, also z.B. ob es sich um ein künstliches neuronales Netz oder einen Entscheidungsbaum (engl. *decision tree*) handelt.¹⁵

II. Historisches Beispiel von KI-Overselling

Der Ansatz zur Übervermarktung von KI ist nicht neu. Eines der frühesten Experimente für Mustererkennung und regelbasierte Antwort-Systeme ist Eliza. Eliza wurde im Jahr 1966 von Joseph Weizenbaum entwickelt und gilt als erster richtiger Chatbot, als Wegbereiter für den Bereich *natural language understanding* und hatte damit massiven Einfluss auf den gesamten NLP-Bereich.

11 C. Weeks, Machine Learning: The Real Buzzword Of 2020. Forbes. 24. Mai 2020, <https://www.forbes.com/sites/forbestechcouncil/2020/03/24/machine-learning-the-real-buzzword-of-2020/?sh=1ec990b63ea7> (besucht am 30.09.2022).

12 T. Fountaine/B. McCarthy/T. Saleh, Building the AI-Powered Organization. Harvard Business Magazine. Aug. 2019, <https://hbr.org/2019/07/building-the-ai-powered-organization> (besucht am 30.09.2022).

13 A. Bibal et al., Legal requirements on explainability in machine learning, in: Artificial Intelligence and Law 29.2 (Juni 2021), 149 (151); <https://link.springer.com/10.1007/s10506-020-09270-4> (besucht am 30.09.2022).

14 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

15 Bibal et al., Explainability, (Fn. 13), 149 (159).

Die Grundfunktionalität von Eliza als virtuelle Therapeutin ist sehr ähnlich zu dem, was man noch heute in Chatbots findet. Gemäß Weizenbaum wird beim Eliza-Bot der Nutzerinput auf bestimmte Keywords untersucht, welche dann anhand von bestimmten Regeln („decomposition rules“) eine möglichst passende Antwort von Eliza hervorrufen.¹⁶ Hierzu hat Weizenbaum zahlreiche mögliche Konversationsverläufe modelliert und die Sprache der Antwort auf die eines echten Therapeuten angepasst (s. Abbildung 1). Für den Endnutzer bedeutet das, dass er seine Fragen und Antworten an Eliza selbst in einer bestimmten Art und Weise formulieren muss, damit Eliza versteht, was gemeint ist.

Ein solches regelbasiertes System bietet einige Herausforderungen. Weizenbaum nennt hierbei sowohl die Erkennung von Keywords und Kontext als auch die Weiterentwicklung des Programms an sich.¹⁷ Somit ist Eliza sehr inflexibel, da sie nur bestimmte Formulierungen erkennen und nicht selbstständig neue Situationen erfassen und verstehen kann. Insofern handelt es sich bei Eliza um eine schwache KI, wenn man sie denn überhaupt „intelligent“ nennen kann.

“[...]”

It's true. I am unhappy.

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

I need some help, that much seems certain.

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

Perhaps I could learn to get along with my mother.

TELL ME MORE ABOUT YOUR FAMILY

My mother takes care of me.

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE

You don't argue with me.

16 Joseph Weizenbaum, ELIZA – a computer program for the study of natural language communication between man and machine, in: Communications of the ACM 9.1 (1966), 36 (38), <https://dl.acm.org/doi/10.1145/365153.365168> (besucht am 29.09.2022).

17 Weizenbaum, ELIZA (Fn. 16).

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

My farther is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE

[...]"

Abbildung 1: Ausschnitt einer Konversation mit Eliza. Regulärer Text: Nutzerinput. Text in Großbuchstaben: Antworten von Eliza.¹⁸

Trotzdem konnte Eliza einige Nutzer von sich überzeugen, insbesondere diejenigen, die dazu instruiert waren, mit Eliza wie mit einem echten Therapeuten zu interagieren.¹⁹ Die Diskussion um diese „Illusion“²⁰ einer Konversation mit einem echten Menschen ist bis heute noch relevant. Zum einen, weil es mit Modellen wie etwa dem Turing-Test bereits Ansätze gibt, welche aktiv hinterfragen, ob eine Maschine selbstständig denken kann.²¹ Zum anderen, weil es mit dem LaMDA-Modell von Google einen aktuellen Fall gibt, bei dem das Bewusstsein von KI diskutiert wurde.

Der Turing-Test wurde im Jahr 1950 vom britischen Logiker, Mathematiker, Kryptoanalytiker und Informatiker Alan Turing entwickelt.²² Im Turing-Test wird überprüft, ob eine Maschine Menschen davon überzeugen kann, dass sie selbst auch ein Mensch ist. Hierzu gibt es drei Testteilnehmer: den *interrogator* (menschlich), die Maschine und einen weiteren Menschen. Der *interrogator* stellt der Maschine und dem Menschen Fragen und bewertet abschließend, welcher Testteilnehmer der Mensch und welcher die Maschine ist. Hierzu muss das KI-System zahlreiche Fähigkeiten aufweisen, wie etwa das *machine learning*, um selbstständig neue Konzepte im Gespräch

18 Joseph Weizenbaum, ELIZA – a computer program for the study of natural language communication between man and machine, in: Communications of the ACM 9.1 (1966), 37 (39), <https://dl.acm.org/doi/10.1145/365153.365168> (besucht am 29.09.2022).

19 Weizenbaum, ELIZA (Fn. 16), 36 (42).

20 Weizenbaum, ELIZA (Fn. 16), 36 (42).

21 G. Oppy/D. Dowe, The Turing Test, Stanford Encyclopedia of Philosophy Archive, 4. Okt. 2021, <https://plato.stanford.edu/archives/win2021/entries/turing-test> (besucht am 30.09.2022).

22 A. Turing, Computing Machinery and Intelligence, in: Mind LIX.236 (Okt. 1950), S. 433 ff.

zu lernen, oder das *natural language processing*, um zu verstehen, was der Gesprächspartner kommuniziert.²³ Ursprünglich als *imitation game* gestartet, wird er bis heute dazu genutzt, um KI-Systeme zu testen und zu bewerten, zum Beispiel beim Loebner Prize.²⁴ Der Turing-Test ist bereits zur Zeit seiner Veröffentlichung mit sechs Einsprüchen (*objectives*) kritisch hinterfragt worden,²⁵ da der Test nicht einwandfrei beweisen kann, dass eine Maschine wirklich denken kann – oder aber dass ein Mensch nicht denkt. Es geht beim Turing-Test vor allem darum, den *interrogator* zu überzeugen. Hierzu stellt sich die Frage, ob dies mit heutigen KI-Systemen und den neuen Sprachmodellen besser funktioniert als mit den logikbasierten Systemen aus der Zeit Turings wie etwa Eliza. Ein gutes Beispiel ist hierfür der Diskurs rund um das LaMDA-Modell von Google. Hierzu schwappte erst kürzlich eine breite Diskussion um den „Eliza Effekt“ bei der Google KI LaMDA hoch,²⁶ bei welcher der Softwareentwickler Blake Lemoine in zahlreichen Blogartikeln berichtete, dass LaMDA ein echtes Bewusstsein besäße.²⁷

Bei LaMDA handelt es sich um ein sehr großes NLP-Modell, welches in Chatbots eingebaut werden kann, ähnlich wie zum Beispiel das BERT-Modell. Der Unterschied zwischen Modellen wie BERT und LaMDA ist, dass LaMDA vor allem auf Grundlage menschlicher Dialoge trainiert wurde.²⁸ Dadurch kann es in einem Chatbot verbaut, viel „menschlicher“ auf Input reagieren und viel eher einen echten Menschen imitieren als Eliza, denn LaMDA kann z.B. spezifische Antworten auf Fragen und auch in der Tonalität Nuancen erzeugen,²⁹ die LaMDAs Antworten überzeugend wirken lassen. Bei Eliza hingegen kommt die Illusion vor allem dadurch zustande, dass Weizenbaum aufgrund seiner persönlichen Erfahrung wusste, wie ein Therapeut

23 *Russell/Norvig*, Artificial intelligence (Fn. 2), S. 2.

24 *D. Powers*, The total Turing test and the Loebner prize, in: Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning – NeMLaP3/CoNLL '98, Sydney, Australia: Association for Computational Linguistics, 1998, 279 (279).

25 *Oppy/Dove*, Turing Test (Fn. 20).

26 *J. Felton*, The Eliza Effect: How A Chatbot Convinced People It Was Real Way Back In The 1960s. 22. Juni 2022, <https://www.iflscience.com/the-eliza-effect-how-a-chatbot-convinced-people-it-was-real-way-back-in-the-1960s-64155> (besucht am 20.09.2022).

27 *C. Vallance*, Google engineer says Lamda AI system may have its own feelings, in: BBC News (13. Juni 2022), <https://www.bbc.com/news/technology-61784011> (besucht am 21.09.2022).

28 *E. Collins/Z. Ghahramani*, LaMDA: our breakthrough conversation technology, <https://blog.google/technology/ai/lamda> (besucht am 21.09.2022).

29 *Collins/Ghahramani*, LaMDA (Fn. 27).

mit seinen Patienten interagiert – und wie erwähnt funktionierte dies vor allem nur dann gut, wenn die Nutzer dazu instruiert waren, sich wie ein Patient zu verhalten. LaMDA ist aufgrund seiner großen Trainingsdatensets viel flexibler.³⁰

Insofern liegt der wesentliche Unterschied zwischen diesen beiden Modellen bzw. Chatbots vor allem in ihrer technischen Entwicklung. Während bei Eliza viele der Regeln und Dialoge von Hand geschrieben und einprogrammiert wurden, haben große IT-Firmen wie Google ganz andere menschliche wie technische Möglichkeiten, um KI-Systeme zu entwickeln. Dadurch kann diese Illusion, also wieder eine Black-Box für den Endnutzer, entstehen, die die Dialoge für den Nutzer so überzeugend machen. Zudem kann man den Sachverhalt rund um LaMDA auch so betrachten, dass diese Berichte bewusst veröffentlicht wurden, um ggf. die Black-Box zu verstärken.

Dennoch bleibt LaMDA ein Modell und ist limitiert in seinen Fähigkeiten, in dem Sinne, wie die Entwickler es designen, genauso wie Eliza. Dabei scheint das Modell für den Nutzer mehr zu können als es eigentlich kann, es wird also erneut „oversell“. Gleichzeitig muss man hier kritisch hinterfragen, ob nicht auch auf Seite der Rezipienten von LaMDA eine Art *overbuying* des KI-Systems gegeben hat. Bei Eliza gab es ein *overbuying*, d.h. das Gefühl von Menschlichkeit war stärker als Weizenbaum es gedacht hatte.³¹ Diese Beobachtung und der erwähnte Diskurs um LaMDA legen nahe, dass es dieses *overbuying* vielleicht auch bei LaMDA gegeben hat.

Aus diesem Grund sollte es nicht nur allgemein diesen Diskurs rund um Black-Boxen, Black-Clouds und die Illusion von KI geben, sondern es sollten auch sowohl Endnutzer, Software-Entwickler als auch Produktmanager aus verschiedenen Bereichen teilnehmen. Gleichzeitig braucht es, wie man am Beispiel vom Turing-Test gesehen hat, Metriken, um überprüfen und testen zu können, wie „bewusst“ KI-Systeme agieren können. Hierzu hat Google nach dem LaMDA-Diskurs angekündigt, an einer Reihe von neuen Tests zu arbeiten, die besser als der bisherige Turing-Test funktionieren sollen und sich dem Konzept der General Artificial Intelligence nähern sollen.³² Das bedeutet, dass hier nicht nur z.B. die Sprachfähigkeit eines KI-Systems getestet wird, sondern auch andere Fähigkeiten wie etwa das Schachspielen.

30 Vallance, Own Feelings (Fn. 26).

31 Weizenbaum, ELIZA (Fn. 16), 36 (42 f.).

32 B. Schwan, Künstliche Intelligenz und LaMDA: Warum Google einen neuen Turing-Test will, heise online, 21. Juni 2022, <https://www.heise.de/hintergrund/Kuenstliche-Intelligenz-Warum-Google-einen-neuen-Turing-Test-will-7145019.html> (besucht am 28.09.2022).

Insofern nähert man sich durch neue Innovation und Gegenentwürfe mit Testmodellen sowie deren Diskurs sehr stark an die Erforschung von KI und Wahrnehmung an. Gleichzeitig ist dies immer noch ein sehr individuelles Unterfangen, wodurch sich viele verschiedene Perspektiven in diesem Diskurs ergeben. In diesem Sinne sollten dabei nicht nur die Datensätze und Grundfunktionalitäten diskutiert werden, sondern auch die Wirkung eines KI-Systems auf den Nutzer. Wir möchten in dem folgenden Kapitel unseren Beitrag zu der Diskussion vertiefen, warum es nicht nur für den Endnutzer, sondern auch für manche IT-Experten KI-Systeme Black-Boxen sind. Im folgenden Kapitel wird dies anhand des Trends des künstlichen neuronalen Netzes und Agenten-Systems aufgezeigt, und deren biologische Vorbilder werden beleuchtet.

C. Technische Komplexität – Neuronale Netzwerke und Datenpipeline

Nachdem wir in den vorherigen Kapiteln einen Blick in die Hintergründe des Entstehens von Black-Boxen geworfen haben, beleuchten wir nun explizit die technische Seite von KI nach aktuellem Forschungsstand. KI kann nämlich nicht nur für Nutzer ohne KI-Hintergrund eine Black-Box und folgend eine Black-Cloud sein, sondern auch für IT-Experten. Dies liegt unserer Ansicht nach vor allem an der Modellierung hinter künstlicher Intelligenz und der (technischen) Komplexität der praktischen Umsetzung dieser Modelle. Dies wird in diesem Kapitel anhand des gegenwärtigen Trends des maschinellen Lernens beleuchtet.

I. Maschinelles Lernen – Grundbegriffe

Maschinelles Lernen (engl. *machine learning*, ML) bezeichnet die Fähigkeit eines KI-Systems, Muster in Daten zu erkennen und darauf basierende Vorhersagen zu treffen und Aktionen zu vollziehen. Grundsätzlich gibt es drei Paradigmen für das maschinelle Lernen:

1. das überwachte Lernen (engl. *supervised learning*),
2. das unüberwachte Lernen (engl. *unsupervised learning*) und
3. das verstärkende Lernen (engl. *reinforcement learning*).

Ein ML-System innerhalb eines der genannten Paradigmen ist dazu designet, ein bestimmtes Problem zu lösen. Dazu nutzt es weitere technische Kompo-

nenten wie z.B. *computer vision*, *natural language processing* oder die Robotik – je nachdem welche Form von Input vorliegt. Ein KI-System besteht also aus mehreren Komponenten, welche miteinander kombiniert und aufeinander angepasst werden müssen.

Typische Probleme und Sachverhalte, mit denen sich die heutigen KI-Systeme in den genannten Paradigmen auseinandersetzen, sind z.B. im überwachtem Lernen die Fähigkeit, eine bestimmte Tierart oder auch Menschen anhand von Trainingsdaten auf einem Bild zu erkennen (Klassifikation) oder die Kaufpreisentwicklung von einem Haus in einer bestimmten Gegend vorherzusagen (Regression). Darüber hinaus kann man mit dem verstärkenden Lernen einer KI etwa das Spielen von Schach beibringen.

Insofern sind die Fähigkeiten der derzeit gängigen ML-Systeme noch eingeschränkt bzw. „schwach“, da KI-Systeme nach heutigem Stand nur ein Problem oder eine eingeschränkte Anzahl an Problemen gleichzeitig lösen können. Eine vollständige Nachahmung einer menschlichen Intelligenz, also einer „starken“ KI, erscheint damit erst mal unwahrscheinlich.

Dennoch steckt hinter diesen vermeintlich „einfachen“ Systemen bereits einiges an Aufwand und interdisziplinärer Arbeit mit Ansätzen zu Fragen der Intentionalität und des Bewusstseins. Hierzu beleuchtet das folgende Kapitel die Umsetzung von KI als Modell von menschlicher und tierischer Intelligenz.

II. KI als Modell – Zwischen Intentionalität und Datenverarbeitung

Ein wichtiger Aspekt von aktuellen KI-Systemen ist die mathematische Erfassung und Modellierung von tierischer wie menschlicher Intelligenz. Die genannten ML-Paradigmen können mithilfe von künstlichen neuronalen Netzwerken (KNN, engl. *artificial neural network*) bearbeitet werden, einer direkten Modellierung der Informationsverarbeitung innerhalb des menschlichen wie tierischen Gehirns. Aufgrund der technologischen Hardware-Entwicklung, insb. von Grafikkarten, hat diese Form des maschinellen Lernens in den letzten 20 Jahren viel Aufmerksamkeit erhalten. Einige der dahinterliegenden Ideen, wie etwa die des künstlichen Neurons, werden jedoch schon seit den 1950er Jahren diskutiert.³³ KI wird hier betrachtet als ein Modell von biologischer Intelligenz (*modelling of* bzw. Modellierung von)

33 *Russell/Norvig*, *Artificial intelligence* (Fn. 2), S. 727 f.

und einer Art Versuch, Intelligenz mathematisch nachzubilden (*modelling for* bzw. Modellierung für).³⁴

Die KI-Forschung ist dabei nicht nur Modellierung, sondern mehr eine Studie und Reflexion unserer selbst, an der viele verschiedene Forscher aktiv teilnehmen, z.B. Biologen, Philosophen und Informatiker. Es ist also wichtig zu verstehen, was die Grundlagen der biologischen Existenz eines intelligenten Wesens sind, ob und wie man diese mathematisch erfassen und für einen Computer greifbar machen kann, sowie das KI-System und seine Nutzung kritisch im Sinne der Philosophie zu hinterfragen.

Um einer Maschine beizubringen, Muster zu erkennen und „selbstständig“ bzw. intentional zu lernen und zu handeln, hat man sich nicht nur intensiv mit der Weitergabe von Informationen bei Menschen und Tieren beschäftigt, sondern auch gezielt mit dem biologischen Lernprozess und der Interaktion zwischen einem Individuum und seiner Umwelt. Hieraus ergeben sich unterschiedliche Komplexitäten in den Anwendungsgebieten von KI-Systemen.

Während KI-Systeme vor allem im *supervised learning* „nur“ große Datenmengen verarbeiten und in ihrem Sichtfeld eingeschränkt sind – durch die Überwachung wird bereits das Ziel des Lernens vorgegeben – so bietet das Paradigma des *unsupervised learnings* und gerade das *reinforcement learning* viel Raum zum Diskurs rund um die Frage der Intentionalität.

Zur Frage der Intentionalität muss man zunächst über den Begriff „Intelligenz“ allgemein sprechen und definieren, was dieser eigentlich meint. Ähnlich wie bei den verschiedenen Paradigmen rund um das maschinelle Lernen gibt es auch beim Intelligenzbegriff verschiedene Unterkategorien, wie etwa soziale oder mathematische Intelligenz. Zudem stellt sich die Frage, ob sich die Intelligenz von Menschen und Tieren grundsätzlich unterscheidet oder ob dieselbe Art von Intelligenz vorliegt. Wie eingangs erwähnt, basieren aktuelle KI-Systeme vor allem auf der Modellierung menschlicher Intelligenz, daher ist hier durchaus Potenzial für weitere Forschung – und damit für weitere Black-Boxen und Black-Clouds. Bei dem Begriff der Intentionalität verhält es sich ähnlich. Hier kann man zwei Grundannahmen treffen:

1. die Intentionalität, welche einem Lebewesen oder Objekt zugeschrieben wird und
2. die Intentionalität, welche das Lebewesen oder Objekt eigentlich hat.

34 C. Geertz, *The interpretation of cultures: selected essays*. ACLS Humanities E-Book. New York: Basic Books, 1973, S. 93 f.

Viele Tierarten – so auch Menschen – besitzen Intentionalität. Maschinen hingegen wirken für viele Menschen so, als hätten sie Intentionalität (vgl. bspw. die LaMDA-Diskussion). Russell und Norvig stellen hierzu die Idee des „rationalen Agenten“ vor, der handelt, um das für sich bestmögliche Ergebnis zu erzielen.³⁵ Diese Idee kann eingeschränkt auch auf KI-Systeme übertragen werden, denn oftmals interagieren sie mit schwierigen Umgebungen.³⁶ Dabei werden sie etwas konkreter mit der Unterscheidung zwischen menschlichem und rationalem Handeln und Denken.³⁷

Auf der einen Seite liegt mit dem Zuschreiben von Rationalität bei Russell und Norvig bereits ein Zuschreiben von Intention vor, denn nach dem bestmöglichen Ergebnis zu handeln ist durchaus eine Intention. Dieser Gedanke ist vor allem im Paradigma des *reinforcement learning* interessant, da hier die Interaktion zwischen einem Individuum, einem Agenten und seiner Umgebung modelliert wird. Anhand von Belohnungen und Bestrafungen soll sich der Agent in seiner Umgebung zurechtfinden und ein Ziel erfüllen. Im Anwendungsfall in der Robotik, mit dem Roboter als Agenten, kann das Ziel sein, sich in einem Raum zu bewegen, ohne gegen ein Hindernis zu stoßen. Durch den Zusammenstoß kann der Roboter lernen, zukünftig Hindernisse besser zu erkennen und nicht mehr mit ihnen zu kollidieren. Hier ist die Intention also, zu lernen sich besser zu bewegen – ähnlich wie bei einem Kleinkind, was gerade das Laufen lernt.

Auf der anderen Seite kann man argumentieren, dass diese Rationalität eine Designentscheidung des Entwicklerteams einer KI ist. Der Grund dafür ist, dass auch diese KI als schwache KI nach wie vor in einem abgesteckten Szenario Anwendung findet. Wenn ein Entwicklerteam einer *reinforcement learning*-KI bspw. das Go spielen beibringt, so ist das vermeintlich rationale Ziel der KI das Go-Spiel gegen jeden Spieler zu gewinnen. Gleichzeitig ist dies aber auch das Ziel des Entwicklerteams, gemessen daran, dass die Sieg-Statistik von Systemen wie AlphaGo eine Metrik dafür ist, wie erfolgreich das AlphaGo-System ist und wie man es weiter verbessern kann.³⁸ Hier stellt sich nachfolgend die Frage, ob das KI-System die Entscheidung über den Spielzug aufgrund des statisch gesehen besten Wertes trifft oder ob es wirklich intentional und intelligent gehandelt hat. Maschinen können daher

35 Russell/Norvig, Artificial intelligence (Fn. 2), S. 4.

36 Russell/Norvig, Artificial intelligence (Fn. 2), S. 4 f.

37 Russell/Norvig, Artificial intelligence (Fn. 2), S. 2.

38 D. Silver et.al., Mastering the game of Go with deep neural networks and tree search, in: Nature 529.7587 (28.01.2016), 484 (484, 487); <https://www.nature.com/articles/nature16961> (besucht am 30.09.2022).

mit Menschen symbiotisch Intentionalität aufweisen, aber die Frage über die Maschinen-Intentionalität ist offen – genauso wie die der eigentlichen Intelligenz der Maschine.

Die Frage nach der Intention und Intelligenz von KI-Systemen ist also noch nicht komplett beantwortet. Hierzu wird sich voraussichtlich in der Forschung rund um die (menschliche) Intelligenz noch einiges tun. Es ist zu erwarten, dass sich die Forschung im Bereich der (menschlichen) Intelligenz noch weiter entwickeln wird. Dies ist ein weiterer wichtiger Grund für die Black Box und ein wichtiger Indikator für ihre weitere Entwicklung und Perspektiven. Im folgenden Kapitel wird das praktische Paradigma der KI beleuchtet und die konkrete technische Umsetzung von KI-Systemen analysiert.

III. KI-Projekte in der Praxis: Teamarbeit und technische Umsetzung

Ein weiterer Grund für die Komplexität von KI-Systemen und somit für die Black-Boxen und Black-Cloud ist die komplexe praktische Umsetzung von KI-Systemen.

KI-Systeme sind datenbasierte Projekte, welche durch die interdisziplinäre Zusammenarbeit von verschiedenen akademischen Feldern geprägt werden. Auch in der Praxis sind, ähnlich wie bei anderen IT-Projekten, meist mehrere Akteure mit unterschiedlichen Hintergründen beteiligt. Die genaue Größe und Zusammensetzung unterscheidet sich meist nach der Unternehmensgröße, es gibt kein perfektes Rezept zur Aufstellung eines KI-Teams. Stobierski schlägt beispielsweise eine Aufteilung in Datenwissenschaftler (eng. *data scientists*), Datenentwickler (engl. *data engineers*) und Datenanalysten (eng. *data analysts*) vor.³⁹ Diese Rollen haben unterschiedliche Aufgaben und Themenschwerpunkte. So kümmert sich der Datenentwickler vor allem um die Infrastruktur der Daten, z.B. um die Sicherung der Daten in Datenbanken, während der Datenwissenschaftler sich mehr mit dem Konzept und den Forschungsfragen zu den vorhandenen Daten auseinandersetzt. Darüber hinaus kann es noch weitere Akteure in dieser Art Projekt geben, wie etwa Projekt- und Produktmanager.

39 T. Stobierski, How to Structure Your Data Analytics Team. Harvard Business School Online. 9. März 2021, <https://online.hbs.edu/blog/post/analytics-team-structure> (besucht am 30.09.2022).

Aus diesen Positionen ergibt sich das folgende Schema zur praktischen Umsetzung datenbasierter Projekte:

1. Erfassen und Vorverarbeiten (engl. *preprocessing*) des Inputs,
2. Analysieren und Verarbeiten des Inputs mithilfe des KI-Systems,
3. Auswerten und Überprüfen des Ergebnisses des KI-Systems,
4. Verbesserung des KI-Systems und
5. Präsentation und Einbindung der Ergebnisse des KI-Systems z.B. in eine Präsentation oder in einen Chatbot.

Dieses Schema ist nicht als linearer Weg zu betrachten, sondern mehr als eine Art Zyklus, bei denen die Schritte ineinander fließen und auch wiederholt werden können, wenn sich z.B. Kundenanforderungen ändern oder man im Entwicklungsprozess neue interessante Erkenntnisse aus den vorliegenden Daten gewinnt. Daher ist für das Schema wichtig, dass die Intention und das Konzept hinter dieser Datenanalyse und praktischen Umsetzung sowie mögliche Forschungsfragen im praktischen Prozess genauso mitwachsen wie die eigentliche Analyse und Umsetzung.

Diese Arbeitsweise ist in der praktischen Umsetzung sehr agil und kann daher sehr gut in Kombination mit Frameworks zur Projektorganisation wie etwa Scrum oder Kanban umgesetzt werden. Gleichzeitig bedeutet dies aber auch mehr Komplexität innerhalb des Projekts durch den Mehraufwand an nötiger Kommunikation in Form von täglichen Meetings für Rücksprachen und offene Fragen.

IV. Technische Umsetzung von KI

Im Folgenden soll die technische Umsetzung eines KI-Systems grob anhand des zuvor erwähnten Schemas erläutert werden. Als konkretes Beispiel dient die Analyse von Social Media Daten.

1. Erfassung der Rohdaten

Grundsätzlich kann man sich nach einer ersten Definition des Konzepts oder einer Forschungsfrage die Frage stellen, ob es bereits einen passenden vorgefertigten Datensatz gibt, z.B. von der Plattform Kaggle oder der Google Data Search, oder ob man selbst manuell die Daten für das Projekt aggregieren muss. Im Fall der Analyse von Social-Media-Daten bietet sich das Crawling von sog. APIs an.

Einige soziale Netzwerke wie Twitter⁴⁰ oder Reddit⁴¹ haben eigene, öffentliche APIs, anhand denen man automatisiert auf bestimmte Daten der Plattformen zugreifen kann. Diese Plattformen bieten Metadaten zu einzelnen Beiträgen an, z.B. den Inhalt des Beitrags, den Autor oder den Veröffentlichungszeitpunkt. Darüber hinaus kann man Kommentare und deren Metadaten zu einem Post herunterladen. Diese Rohdaten können über einen bestimmten Zeitraum hinweg gemäß der Limitationen der API gesammelt und analysiert werden.

An dieser Stelle ist zu hinterfragen, ab welcher Größe der Datensatz für die Analyse relevant wird. Eine konkrete Antwort gibt es hierauf nicht, allerdings kann man schlussfolgern: Je größer und diverser ein Datensatz aufgestellt ist, desto besser. Im Sinne des Big Data-Gedankens leben die meisten Werkzeuge der Datenwissenschaft bzw. *data science* von Statistik – so auch das maschinelle Lernen. Daher macht es Sinn, so viele hochwertige Daten wie möglich zu sammeln. Dies ist auch notwendig, um genügend Daten für das Trainieren, Evaluieren und Testen eines KNNs zu haben, um die KNN-Performance korrekt beurteilen zu können.

2. Vorbereitung und Analyse der Daten

Hat man die notwendige Datenmenge zusammen, kann man bereits eine erste Analyse der Rohdaten vollziehen und die passende Analysemethode auswählen. Im Fall von Social Media-Daten kann dies z.B. Text sein. Die textuellen Social Media-Beiträge kann man mithilfe von bspw. Textklassifikation mit KNNs wie etwa BERT oder auch durch eine statische Analyse der Postinhalte mit *topic modelling* untersuchen. Hierfür müssen die Rohdatensätze etwas gesäubert und vorbereitet werden, z.B. durch die Lemmatisierung der Textinhalte aus den Social Media-Beiträgen.

Die Analyse von Daten mit künstlichen neuronalen Netzwerken ist dabei sehr komplex. Wie bei den Datensätzen hat man auch hier die Möglichkeit ein bereits vorgefertigtes Modell zu wählen, z.B. ImageNet für Bildklassifikation oder BERT für Textklassifikation, oder selbst ein KNN zu entwickeln.

40 Inc. Twitter. Twitter API Documentation | Docs | Twitter Developer Platform. 2022, <https://developer.twitter.com/en/docs/twitter-api> (besucht am 30.09.2022).

41 J. Wardle, API, 2022, <https://github.com/reddit-archive/reddit/wiki/API> (besucht am 30.09.2022).

V. Komplexität der praktischen Umsetzung von KNNs

Wie bereits in Kapitel unter C. II. dargelegt sind KNNs eine Modellierung der biologischen Informationsverarbeitung von Menschen und Tieren. In der Praxis kann dies über sog. *multilayer feed-forward neural networks* umgesetzt werden (s. Abbildung 2).

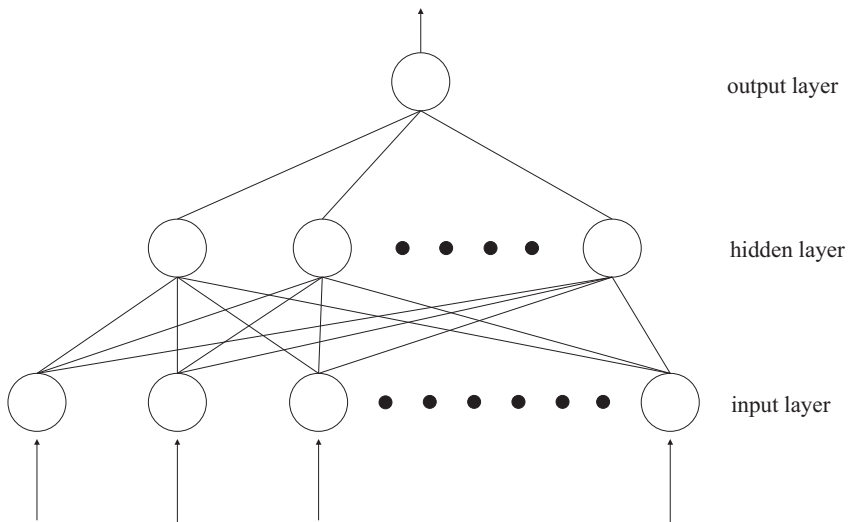


Abbildung 2: Schema eines multilayer feed-forward neural networks

Dieses Modell nimmt in den Input-Ebenen Input in Form von z.B. einem Text an und verarbeitet in versteckten Ebenen (engl. *hidden layers*) bestimmte Attribute dieses Inputs. In der *feature extraction* wird der Text-Input in der Input-Ebene in eine Form gebracht, mit welcher das KNN die Klassifikation durchführen kann. Hierzu wird mit Modellen wie etwa dem *bag-of-words* jeder Satz in einen Vektor umgewandelt, bei dem jede Stelle ein Wort impliziert. Durch den Abgleich und die Verrechnung der daraus generierten Features kann das KNN im Fall einer Textklassifikation dann vorhersagen, in welche Kategorie der Text gehört. Dies geschieht in den versteckten Ebenen. Die Output-Ebene führt die Ergebnisse dann in eine abschließende Vorhersage-Statistik zusammen.

Die künstlichen Neuronen des KNNs sind miteinander verbunden. Die Informationsverarbeitung läuft dabei linear von links nach rechts, durch die Verbindungen zwischen den Ebenen kann sich das KNN über das

Konzept der *backpropagation* im Trainingsvorgang selbst verbessern. Hier wird mithilfe einer Verlustfunktion bereits im Training das Zwischenergebnis der Klassifikation überprüft, und die Parameter des KNNs werden angepasst. Diese Parameter, z.B. die Gewichtung bestimmter Informationen, sind für die korrekte Klassifikation des Textes relevant. Ob eine Information anhand ihrer Gewichtung von einem künstlichen Neuron zum nächsten Neuron weitergegeben wird, ist davon abhängig, ob die Aktivierungsfunktion aktiviert werden kann. Das bedeutet, dass man initial als Entwickler die Gewichtungen und weitere Parameter wie die *learning rate* vorgibt, anhand derer das KNN sich sozusagen selbst verbessert und die Klassifikation berechnet. Die Berechnung der Anpassung der Gewichte des Netzes erfolgt selbstständig – man kann dies am Anfang als Entwickler einstellen und nach erfolgreichem Training nachjustieren. Darüber hinaus gibt es weitere Mechanismen wie etwa die Regularisierung (engl. *regularization*), die man einbauen kann, um das *overfitting* eines Modells, also das Auswendiglernen zu verhindern. Bei der Regularisierung werden z.B. einige Parameter des Modells bewusst weggelassen, um mehr Übersicht zu gewinnen. Wichtig ist zu betonen, dass all diese Bausteine je nach Problem unterschiedlich ausfallen können. So gibt es z.B. mehrere Arten von Aktivierungsfunktionen oder auch Verlustfunktionen. Sobald man die Komponenten richtig zusammengebaut und den Trainingsvorgang gestartet hat, kann man die Parameterweitergabe und Klassifikation nicht mehr beeinflussen bis man einen zweiten Trainingsvorgang startet. Insofern gibt es hier eine Black-Box by design, da man so das selbstständige Lernen des KNNs garantieren will.

Zusammenfassend ist die praktische Umsetzung von KI-Systemen ein komplexes und vielschichtiges Unterfangen, bei dem nicht umsonst viele verschiedene Akteure mit unterschiedlichen Aufgabengebieten in KI-Projekten teilnehmen. Neben der Black-Box by design bei KNNs als Analyseverfahren ist es ebenso wichtig, über die Erstellung der Datensätze und deren technische Umsetzung zu sprechen. Nur wenn ein umfassender und diverser Datensatz verfügbar ist, können qualitativ hochwertige Vorhersagen mit Black-Box-Systemen wie KNNs gemacht werden, wo ein Entwickler nicht in jeden Entscheidungsprozess des KNNs hineinschauen und nachjustieren kann. Erwähnt sei hier ebenfalls, dass es für diesen Diskurs ebenso wichtig ist, auch über die Präsentation und Visualisierung der Ergebnisse, z.B. in Form einer Konfusionsmatrix, zu sprechen, da Visualisierungen genauso Bias aufweisen und die Interpretation von KI beeinflussen können wie die eigentliche Analyseverfahren selbst.

D. Diskussion

Die vorangegangenen Kapitel haben gezeigt, wie vielfältig der Diskurs rund um KI-Systeme und ihre Funktionalität ist. Egal ob IT-Experte oder Laie, jeder bringt eine individuelle Perspektive in den Diskurs ein, abhängig von seinem Hintergrundwissen und (täglichen) Umgang mit KI-Systemen.

Um dieses individuelle Verständnis akkurat darzustellen, möchten wir nachfolgend die Black-Cloud-Formel vorstellen und einige Argumente dieses Aufsatzes kritisch im Ausblick hinterfragen.

I. Black-Cloud-Formel

Die vorliegende Formel soll das individuelle Verständnis einer Person von KI-Systemen mathematisch modellieren. Die Schwärze s jedes Punkts p in der Black-Cloud BC ist Resultat einer Funktion mit Vorkenntnissen (auch Kompetenz und Hintergrund) V und der technischen Komplexität K des KI-Systems als Parametern. D.h.

$$s_p = f(V, K)$$

und BC ist die Menge von allen p mit einem Wert höher als einem Grenzwert g .

$$p \in BC \Leftrightarrow s_p > g$$

Dadurch soll in diesem Modell die dynamische Entwicklung des Verständnisses dargestellt werden, durch welche die Black-Cloud flexibler ist und sich mit dem Untersuchungsobjekt verändert.

II. Ausblick

Nach einer ersten Beurteilung der Nutzer von KI-Systemen wäre es weiterhin interessant zu untersuchen, wie sich das Verständnis von KI-Systemen auf kurz- und langfristiger Basis verändert. Hierbei ist kritisch anzumerken, dass in einer Diskussion um KI-Systeme ebenfalls berücksichtigt werden sollte, ob sich die verschiedenen Nutzergruppen überhaupt für die Funktionalität von KI-Systemen interessieren oder ob sie diese Produkte ohne Hintergedanken nutzen wollen.

Außerdem ist zukünftig zu untersuchen, wie man die Funktionalität von KI-Systemen am besten erklärt. Wie im Kapitel *Technische Umsetzung von KI* aufgezeigt, besteht diese Art von Produkt aus mehreren Komponenten, an denen komplette Teams arbeiten. Es ist schwierig, dies alles innerhalb eines Beitrags zusammenzufassen – hier kann man noch viel mehr in die Tiefe gehen. Wir haben in dieser schematischen Erklärung beispielsweise viele mathematische Komponenten ausgelassen, die für manche Akteure des KI-Diskurses von größter Wichtigkeit sind.

E. Zusammenfassung

Zusammenfassend zeigt dieser Beitrag auf, wie wichtig und vielfältig der Diskurs rund um KI-Systeme ist. Die Diskussionen rund um das Eliza-, aber auch um das LamDA-Modell zeigen, wie viele verschiedene Perspektiven es gibt – und wie wichtig es ist, diese im Sinne des *overbuying* und *overselling* kritisch zu hinterfragen. Gleichzeitig gibt es bereits erste Untersuchungen, welche erfassen, wer wie viel über KI-Systeme weiß und auch wie man die Sichtbarkeit von KI-Systemen schematisieren kann.

Aus diesem Grund schlagen wir mit der Black-Cloud und ihrer Formel ein erstes Modell vor, um das individuelle Verständnis von KI-Systemen greifbar zu machen. Für die Black-Cloud ist dabei nicht nur das individuelle Vorwissen ein maßgeblicher Faktor, sondern auch die technische Komplexität. Die Diskussion rund um das *overselling* zeigt, dass aus juristischer Sicht in der Wirtschaft KI nicht erklärt werden muss. Dennoch ist KI, z.B. in Form von Machine Learning, eine komplexe Technologie, die ebenso für die Größe der eigenen Black-Cloud wichtig ist. Nicht nur, weil an der Entwicklung solcher Systeme viele verschiedene Akteure beteiligt sind und KI immer eine kollektive Teamleistung ist, sondern auch, weil mit der Modellierung biologischer Intelligenz manchen Systemen eine Art Intention zugeschrieben werden kann und so eine Black-Box by design entsteht. Gleichzeitig gibt es aber auch Argumente, die gegen eine Intentionalität von KI-Systemen sprechen. So können aktuelle KI-Systeme noch immer nur bestimmte Aufgaben erfüllen und sind inflexibel. Doch bereits hier stecken viele verschiedene Bausteine in solchen Systemen, wie das Beispiel von künstlichen neuronalen Netzwerken zeigt.

Es ist abzuwarten, wie sich diese Technologie bzw. dieses Forschungsfeld in den nächsten Jahren weiterentwickeln wird. Bis dahin bietet die Black-Cloud

eine Anregung zum Individualisieren des KI-Diskurses und zur Modellierung des Verständnisses von KI-Systemen.

Völkerrechtliche Schwierigkeiten einer effektiven Bekämpfung von Cyber-Angriffen

Stephan Hobe, Martin Schwaborn

Einleitung

Dass die vielfach verwendete Cybertechnologie nicht nur neue Chancen eröffnet, indem sie etwa autonomes Fahren, Fliegen und Ähnliches ermöglicht, ist spätestens klar geworden, seitdem es Cyberangriffe auf Estland im Jahre 2007, auf Georgien im Jahre 2008, auf den Iran im Jahre 2010,¹ auf den Deutschen Bundestag im Jahre 2017² sowie auf das Universitätsklinikum in Düsseldorf im Jahre 2020³ gegeben hat. Auch der völkerrechtswidrige Angriffskrieg Russlands gegen die Ukraine⁴ wird zwar in erster Linie konventionell geführt, gleichzeitig aber von Anfang an durch umfassende Cyberoperationen flankiert.⁵ Schon vor Beginn des russischen Überfalls auf die Ukraine haben sämtliche Großmächte Cyberspace nicht nur als neue Dimension, sondern auch als neue Domäne der Kriegführung betrachtet. Viele Staaten, unter anderem die USA, Russland, China, das Vereinigte Königreich, Israel, Deutschland, aber auch Nordkorea, haben eigene Cybereinheiten ins Leben

-
- 1 Zu diesen und weiteren Beispielfällen siehe *J. Dornbusch*, Das Kampfführungsrecht im internationalen Cyberkrieg, Baden-Baden 2018, S. 31 ff. sowie *C. Focarelli*, Self-defence in cyberspace, in: N. Tsagourias/R. Buchan (Hrsg.), *Research Handbook on International Law and Cyberspace*, 2. Aufl., Sheffield 2021, S. 317, 321 ff. sowie *J.-C. Woltag*, *Cyber Warfare, Military Cross-Border Computer Network Operations under International Law*, Cambridge (u.a.) 2014, S. 47 ff.
 - 2 „Erneuter Hackerangriff auf Rechner von Abgeordneten“, *Zeit Online* (29.03.2017), abrufbar unter: https://www.zeit.de/politik/deutschland/2017-03/bundestag-hackerangriff-verfassungsschutz-netzsicherheit?utm_referrer=https%3A%2F%2Fwww.google.com%2F (zuletzt abgerufen: 23.02.2023).
 - 3 Pressemitteilung des Universitätsklinikums Düsseldorf vom 17.09.2020, abrufbar unter: <https://www.uniklinik-duesseldorf.de/ueber-uns/pressemitteilungen/detail/it-ausfall-an-der-uniklinik-duesseldorf> (zuletzt abgerufen: 23.02.2023).
 - 4 Allgemein siehe nur *S. Schmahl*, Völker- und europarechtliche Implikationen des Angriffskriegs auf die Ukraine, *NJW* 2022, 969.
 - 5 Einen aktuellen Überblick bietet die Zeitleiste des CyberPeace Institute, abrufbar unter: <https://cyberconflicts.cyberpeaceinstitute.org/threats/timeline> (zuletzt abgerufen: 23.02.2023).

gerufen, die sich vor allem mit der Abwehr entsprechender Angriffe befassen. Allein für Deutschland sei nur die Errichtung des Nationalen Cyber-Abwehrzentrums im Jahre 2011 sowie die Aufstellung des Organisationsbereichs Cyber- und Informationsraum (CIR) der Bundeswehr im Jahre 2017 mit seinem Kommando Cyber- und Informationsraum in Bonn genannt.⁶ Infolge des Ukraine-Krieges wurde jüngst von Bundesinnenministerin *Nancy Faeser* auch eine Debatte über eine Grundgesetzänderung angestoßen, die dem Bund zusätzliche Kompetenzen zur Abwehr von Cyberattacken inklusive der Möglichkeit aktiver Gegenangriffe, sogenannter „Hackbacks“, einräumen soll.⁷

Mit den verschiedenen Formen von Attacken auf Computerbetriebssysteme durch Viren, Würmer, Trojanische Pferde, Logische Bomben, Backdoor oder Denial of Service-Attacken ist die Variationsbreite möglicher Angriffe enorm.⁸ Als einführendes Beispiel für Auswirkungen und Probleme, die mit Cyberangriffen sowohl in faktischer als auch in völkerrechtlicher Hinsicht verbunden sind, soll an dieser Stelle der Computerwurm Stuxnet dienen. Bei Stuxnet handelt es sich um ein Schadprogramm, welches gezielte Angriffe auf ein bestimmtes System zur Überwachung und Steuerung technischer Prozesse ermöglicht.⁹ Genau dieses SCADA-System (aus dem Englischen, Supervisory Control and Data Acquisition), die Simatic S7, wurde unter anderem auch zur Steuerung von Zentrifugen zur Urananreicherung in der iranischen Atomanlage Natanz verwendet. Stuxnet manipulierte die Rotationsgeschwindigkeit der hochsensiblen Zentrifugen, was die iranische Urananreicherung empfindlich gestört und langfristig zu einer Zerstörung von etwa 1000 der 9000 Zentrifugen in Natanz geführt hat. Wer den Computerwurm in die iranische Atomanlage eingeschleust hat, ist nach wie vor

6 Dazu <https://www.bundeswehr.de/de/organisation/cyber-und-informationsraum> (zuletzt abgerufen: 23.02.2023); im Überblick zu anderen Ländern siehe *S.-H. Schulze*, Cyber-„War“ – Testfall der Staatenverantwortlichkeit, Tübingen 2015, S. 14 f.

7 *D. Neuerer*, Ukraine-Krieg: Regierungspläne für Cyber-Gegenangriffe stoßen auf Ablehnung, Handelsblatt vom 04.04.2022, abrufbar unter: https://www.handelsblatt.com/politik/deutschland/hackerattacken-ukraine-krieg-regierungsplaene-fuer-cyber-gegenangriffe-stossen-auf-ablehnung/28225644.html?nlayer=Themen_11804704, (zuletzt abgerufen: 23.02.2023). Allgemein zu Hackbacks von Staaten *H. Lahmann*, “Hacking Back” by States and the Uneasy Place of Necessity within the Rule of Law, ZaöRV (80) 2020, 453.

8 Zu den verschiedenen Angriffsformen siehe *Dornbusch*, Kampfführungsrecht (Fn. 1), S. 30 f. sowie *Schulze*, Cyber-„War“ (Fn. 6), S. 24 f.

9 Zum Stuxnet-Fall siehe nur *Woltag*, Cyber Warfare (Fn. 1), S. 47 ff.

nicht endgültig geklärt. Zwar gehen Experten wegen der Komplexität der Schadsoftware davon aus, dass die handelnden Personen Unterstützung von staatlicher Seite hatten, wobei insbesondere die USA und Israel genannt werden, doch ist die Verantwortung bis heute offen.

Der Stuxnet-Fall weist damit zwei Charakteristiken auf, die für Cyberangriffe typisch sind. Zum einen kann ein vermeintlich harmloses Computerprogramm große Schäden verursachen, was zwangsläufig zur Frage nach möglichen Gegenmaßnahmen führt. Zum anderen ist oftmals nicht oder jedenfalls nicht innerhalb kurzer Zeit zu ermitteln, von wo und vor allem durch wen der Angriff verübt wurde. Vergleichbare Szenarien wie im Stuxnet-Fall, also Angriffe auf kritische Infrastruktur wie Strom- und Wasserversorgung, Krankenhäuser oder Telekommunikationseinrichtungen können jederzeit auch die Bundesrepublik Deutschland oder einen NATO-Bündnispartner treffen.¹⁰ Infolge des Ukraine-Krieges geht das Bundesamt für Sicherheit in der Informationstechnik (BSI) zurzeit von einer „erhöhte[n] Bedrohungslage für Deutschland“ aus,¹¹ weshalb sich die Frage nach der völkerrechtlichen Einordnung derartiger Cyberangriffe besonders deutlich aktualisiert.

Ein besonderer Fokus soll hier auf dem Recht der Selbstverteidigung liegen. Zur besseren Einordnung ist der Beitrag in zwei Teile gegliedert. Den Anfang macht ein allgemeiner Überblick zu den rechtlichen Grundlagen und Voraussetzungen des Selbstverteidigungsrechts souveräner Staaten, einschließlich seiner Stellung im allgemeinen Völkerrecht (A.). Anschließend werden die Herausforderungen für das Selbstverteidigungsrecht beleuchtet, die sich aus den Charakteristiken des Cyberspace und der dortigen Angriffsformen ergeben (B.).

10 Vgl. BMVg, Broschüre der Bundeswehr zu Landes- und Bündnisverteidigung, Juli 2020, S. 14, abrufbar unter: <https://www.bundeswehr.de/resource/blob/2338734/8bcff03f523a3962a028ef20484f3f0b/download-broschuere-de-data.pdf>; BMI, Cybersicherheitsstrategie für Deutschland, 2016, S. 38 ff., abrufbar unter: <https://www.bundeswehr.de/resource/blob/89756/6b2dcb8af248db01ea3e338d8a54e8bb/cybersicherheitsstrategie-data.pdf>; zur Cyberabwehr Deutschlands in der NATO siehe BMVg, Weißbuch 2016 zur Sicherheitspolitik und zur Zukunft der Bundeswehr, S. 65, abrufbar unter: <https://www.bmvg.de/resource/blob/13708/015be272f8c0098f1537a491676bfc31/weissbuch2016-barrierefrei-data.pdf> (alle Links zuletzt abgerufen: 23.02.2023).

11 BSI, Einschätzung der aktuellen Cyber-Sicherheitslage in Deutschland nach dem russischen Angriff auf die Ukraine, Pressemitteilung vom 12.05.2022, abrufbar unter: https://www.bsi.bund.de/DE/Service-Navi/Presse/Pressemitteilungen/Presse2022/220225_Angriff-Ukraine-Statement.html (zuletzt abgerufen: 23.02.2023).

A. Grundlagen und Einordnung des Selbstverteidigungsrechts souveräner Staaten nach Art. 51 UN-Charta

Um die Besonderheiten und Herausforderungen für das Selbstverteidigungsrecht gegenüber Cyberangriffen vollständig erfassen zu können, kommt man nicht umhin, die völkerrechtlichen Grundlagen des in Art. 51 UN-Charta verankerten Selbstverteidigungsrechts zu betrachten. Ausgangspunkt ist Art. 2 Ziff. 1 UN-Charta, wonach die Organisation der Vereinten Nationen auf dem Grundsatz der souveränen Gleichheit aller ihrer Mitglieder beruht. Die sogenannte *Friendly Relations Declaration* der UN-Generalversammlung vom 24. Oktober 1970 hat diesen „Grundsatz der souveränen Gleichheit der Staaten“ wie folgt konkretisiert:

„Alle Staaten genießen souveräne Gleichheit. Sie haben gleiche Rechte und Pflichten und sind ungeachtet wirtschaftlicher, sozialer, politischer oder anderer Unterschiede gleichberechtigte Mitglieder der internationalen Gemeinschaft.“¹²

Abgesichert wird die souveräne Gleichheit aller Mitgliedstaaten der UN durch verschiedene völkerrechtliche Grundprinzipien, die weitgehend auch in Art. 2 UN-Charta festgehalten sind. Für die hier relevante Frage nach dem Umgang mit *Cyberangriffen* ist neben dem sogenannten Interventionsverbot, auf das noch zurückzukommen sein wird, insbesondere das Gewaltverbot von Interesse.¹³ Nach Art. 2 Ziff. 4 UN-Charta unterlassen alle Mitglieder in ihren internationalen Beziehungen jede gegen die territoriale Unversehrtheit oder die politische Unabhängigkeit eines Staates gerichtete oder sonst mit den Zielen der Vereinten Nationen unvereinbare Androhung oder Anwendung von Gewalt. Entsprechend seiner Funktion zur Sicherung des Welt-

12 UN Generalversammlung, Erklärung über Grundsätze des Völkerrechts betreffend freundschaftlicher Beziehungen und Zusammenarbeit zwischen den Staaten im Einklang mit der Charta der Vereinten Nationen, A/RES/2625 (XXV) vom 24.10.1970; zur rechtlichen Einordnung siehe nur *H. Keller*, *Friendly Relations Declaration*, MPEPIL, abrufbar unter <https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e938?rskey=2SnJUN&result=1&prd=MPEIL> (zuletzt abgerufen 23.02.2023).

13 Zu den Grundprinzipien der zwischenstaatlichen Beziehungen siehe nur *M. Herdegen*, *Völkerrecht*, 20. Aufl., München 2021, S. 267 ff.; *S. Hobe*, Einführung in das Völkerrecht, II. Aufl., Tübingen 2020, S. 205 ff.; allgemein zur Einordnung von Cyberoperationen als Verstoß gegen das Gewaltverbot *M. Roscini*, *Cyber operations as a use of force*, in: N. Tsagourias/R. Buchan (Hrsg.), *Research Handbook on International Law and Cyberspace*, 2. Aufl., Cheltenham 2021, S. 297 (301 ff.).

friedens untersagt Art. 2 Ziff. 4 UN-Charta jede Form der Gewaltanwendung. Ohne bereits an dieser Stelle ins Detail gehen zu wollen, können vom Gewaltverbot neben dem klassischen Anwendungsfall der militärischen Gewalt im Sinne einer Waffenanwendung gegen das Hoheitsgebiet oder die Streitkräfte eines anderen Staates auch Maßnahmen unterhalb der Schwelle eines Krieges, wie beispielsweise die Entsendung bewaffneter Rebellen Gruppen oder Banden, fallen.¹⁴ Auf diese letzte Fallgruppe und den in diesem Zusammenhang relevanten *Nicaragua*-Fall des Internationalen Gerichtshofs¹⁵ wird noch im späteren Verlauf des Beitrags zurückzukommen sein.

An dieser Stelle genügt es zunächst einmal festzuhalten, dass Art. 2 Ziff. 4 UN-Charta jede Form der Gewaltanwendung zwischen den Staaten untersagt und dass die UN-Charta nur zwei Ausnahmen von diesem Grundsatz kennt.¹⁶ Ausdrücklich vom Gewaltverbot ausgenommen sind zum einen Maßnahmen der kollektiven Sicherheit nach Art. 39 bis 50 UN-Charta, also Handlungen des UN-Sicherheitsrats zur Friedenssicherung. Die zweite Ausnahme bildet das Selbstverteidigungsrecht der Staaten. Der insoweit relevante Art. 51 UN-Charta beinhaltet folgende Regelung:

„Diese Charta beeinträchtigt im Falle eines bewaffneten Angriffs gegen ein Mitglied der Vereinten Nationen keineswegs das naturgegebene Recht zur individuellen oder kollektiven Selbstverteidigung, bis der Sicherheitsrat die zur Wahrung des Weltfriedens und der internationalen Sicherheit erforderlichen Maßnahmen getroffen hat. Maßnahmen, die ein Mitglied in Ausübung dieses Selbstverteidigungsrechts trifft, sind dem Sicherheitsrat sofort anzuzeigen; sie berühren in keiner Weise dessen auf dieser Charta beruhende Befugnis und Pflicht, jederzeit die Maßnahmen zu treffen, die er zur Wahrung oder Wiederherstellung des Weltfriedens und der internationalen Sicherheit für erforderlich hält.“

Bereits dem Wortlaut von Art. 51 UN-Charta lassen sich verschiedene Anforderungen und Grenzen des staatlichen Selbstverteidigungsrechts entneh-

14 Dazu *Hobe*, Völkerrecht (Fn. 13), S. 207 f.; ausf. zum Gewaltbegriff des Art. 2 Ziff. 4 UN-Charta auch *A. Randelzhofer/O. Dörr* in: *Simma/Khan/Nolte/Paulus, The Charter of the United Nations, A Commentary*, 3. Aufl., Oxford 2012, Art. 2 (4) Rn. 14 ff.

15 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, insb. § 195.

16 Der Ausnahme über die sog. Feindstaatenklausel nach Art. 53 und 107 UN-Charta kommt nach einhelliger Auffassung keine Bedeutung mehr zu. Dazu nur *Randelzhofer/Dörr* in: *Simma/Khan/Nolte/Paulus* (Fn. 14), Art. 2 (4) Rn. 45 m.w.N.

men.¹⁷ Nicht zuletzt, um Wiederholungen im zweiten Teil des Beitrags zu vermeiden, geht es an dieser Stelle nur darum, einen Überblick zu den einzelnen Voraussetzungen und den damit verbundenen Problemen zu geben, die sich bei der Reaktion auf Cyberattacken stellen.

Ausgangspunkt und Grundvoraussetzung des Selbstverteidigungsrechts ist das Vorliegen eines „bewaffneten Angriffs“ auf einen Mitgliedstaat der Vereinten Nationen. Während ein konventioneller, das heißt in diesem Zusammenhang vor allem analoger Militärschlag eines Staates gegen das Territorium oder die Streitkräfte eines anderen Staates regelmäßig ohne Weiteres als „bewaffneter Angriff“ eingeordnet werden kann, ist die Lage bei Cyberattacken, die jedenfalls im Ausgangspunkt im rein digitalen Raum ablaufen, nicht immer eindeutig. Hinzu kommt, dass der „bewaffnete Angriff“ von einem anderen Staat ausgehen oder diesem jedenfalls zurechenbar sein muss. Stellt die Frage der Zurechnung von Handlungen privater Akteure an Staaten das Völkerrecht in Zeiten des internationalen Terrorismus schon bei analogen Angriffen vor nicht unerhebliche Herausforderungen, kommt bei Cyberattacken das Problem der schwierigen – wenn nicht sogar unmöglichen – technischen Rückverfolgbarkeit hinzu. Mit anderen Worten: Der Verursacher kann regelmäßig nicht ermittelt werden. Auch die weiteren Voraussetzungen bzw. die Grenzen des völkerrechtlichen Selbstverteidigungsrechts können im Einzelfall mit nicht unerheblichen Schwierigkeiten verbunden sein. So muss die Selbstverteidigung der Abwehr eines *gegenwärtigen* Angriffs dienen und die ergriffenen Verteidigungsmaßnahmen müssen verhältnismäßig sein. Insbesondere wenn man die schwierige und unter Umständen langwierige Rückverfolgung von Cyberattacken betrachtet, kann das Kriterium der Gegenwärtigkeit des Angriffs bzw. der Unmittelbarkeit der Verteidigung problematisch sein. Unter dem Gesichtspunkt der Verhältnismäßigkeit ist insbesondere zu diskutieren, welche Reaktionen auf Cyberattacken möglich und vor allem angemessen sind. Neben diesen quantitativen und qualitativen Anforderungen an das Selbstverteidigungsrecht kommt noch eine weitere zeitliche Grenze hinzu. So muss die Selbstverteidigung sofort dem UN-Sicherheitsrat angezeigt werden und das Recht zur Selbstverteidigung endet spätestens, wenn der Sicherheitsrat die zur Wahrung des Weltfriedens und der internationalen Sicherheit erforderlichen Maßnahmen getroffen hat. In der Anzeigepflicht kommt die rechtliche Subsidiarität des Selbstverteidi-

17 Zu den Anforderungen siehe nur A. Randelzhofer/G. Nolte, in: Simma/Khan/Nolte/ Paulus (Fn. 14), Art. 51 Rn. 16 ff. m.w.N. sowie sogleich ausf. unter B. II.

gungsrechts gegenüber Maßnahmen des UN-Sicherheitsrats zur kollektiven Sicherheit zum Ausdruck.

Bevor im zweiten Teil des Beitrags die einzelnen Voraussetzungen und Grenzen ausführlicher erläutert und hinsichtlich ihrer Anwendbarkeit auf Cyberattacken einer kritischen Betrachtung unterzogen werden, gilt es im allgemeinen Teil mit dem Interventionsverbot noch ein weiteres völkerrechtliches Grundprinzip für die zwischenstaatlichen Beziehungen zu behandeln. Wie soeben geschildert, ist die Grundvoraussetzung des Selbstverteidigungsrechts das Vorliegen eines „bewaffneten Angriffs“. Wie zu zeigen sein wird, sind gerade im Bereich des Cyberspace nicht wenige Szenarien denkbar, die man umgangssprachlich als „Hacker-Angriff“ einordnen würde, die im juristischen Sinne aber nicht die Schwelle des „bewaffneten Angriffs“ nach Art. 51 UN-Charta bzw. die Schwelle des Gewaltverbots aus Art. 2 Ziff. 4 UN-Charta überschreiten. Dies schließt zwar die Möglichkeit zur Selbstverteidigung aus, stellt den betroffenen Staat aber nicht vollkommen schutzlos. Es kommt nämlich nach wie vor ein Verstoß gegen das völkerrechtliche Interventionsverbot in Betracht.

Das Interventionsverbot schließt eine Einmischung in innere Angelegenheiten eines anderen Staates aus. So untersagt Art. 2 Ziff. 7 UN-Charta, außer im Falle von Maßnahmen der kollektiven Sicherheit, ein Eingreifen der Vereinten Nationen in „Angelegenheiten, die ihrem Wesen nach zur inneren Zuständigkeit eines Staates gehören“. Doch nicht nur für die Vereinten Nationen ist ein solcher Eingriff ausgeschlossen. Auch zwischen den Mitgliedstaaten besteht ein allgemeines Interventionsverbot. Der Internationale Gerichtshof hat im *Nicaragua*-Fall die Geltung des Interventionsverbots als Satz des Völkergewohnheitsrechts unter Bezugnahme auf die bereits angesprochene *Friendly Relations Declaration* ausdrücklich anerkannt.¹⁸ Eine verbotene Intervention setzt voraus, dass eine Einmischung in die inneren Angelegenheiten eines Staates vorliegt, die unter Androhung oder Anwendung von Zwang erfolgt ist.¹⁹ Als innere Angelegenheiten gelten jene Bereiche, die dem Staat vorbehalten sind (sog. *domaine réservé*), wozu der Internationale Gerichtshof die Wahl des politischen, wirtschaftlichen, sozialen und kulturellen

18 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 202.

19 Ausführlich zum Interventionsverbot *U. Scheuner*, *Intervention und Interventionsverbot*, Zeitschrift für die Vereinten Nationen 5/1980, 149.

Systems sowie die Formulierung der auswärtigen Politik gezählt hat.²⁰ Problematischer ist demgegenüber, wann eine bestimmte Handlung als verbotener Zwang einzuordnen ist. Dass eine Cyberattacke, die keinen „bewaffneten Angriff“ im Sinne des Art. 51 UN-Charta darstellt, einen Verstoß gegen das Interventionsverbot begründen kann, ist allerdings alles andere als von vornherein ausgeschlossen. Natürlich stellen sich auch hier die bereits angesprochenen Probleme der Rückverfolgbarkeit und Zurechnung sowie der Angemessenheit einer staatlichen Gegenmaßnahme. Dennoch bietet auch das Interventionsverbot einen völkerrechtlichen Anknüpfungspunkt für einen Umgang mit Cyberattacken, weshalb im weiteren Verlauf des Beitrags auch darauf noch ausführlicher einzugehen sein wird.

B. Cyberangriffe als Herausforderung für das Selbstverteidigungsrecht

Im Anschluss an den kurzen Überblick über die völkerrechtlichen Grundlagen des Selbstverteidigungsrechts können nun die Herausforderungen in den Blick genommen werden, die mit den schier unbegrenzten Angriffsmöglichkeiten im Cyberspace und ihren Auswirkungen einhergehen. Um diese nicht nur zu erfassen, sondern möglichst auch völkerrechtlich zu bewältigen, muss zunächst eine wichtige Vorfrage geklärt werden. Ausgangspunkt der Überlegungen muss selbstverständlich die Frage sein, was unter Cyberspace und Cyberattacken zu verstehen ist und vor allem, ob und wenn ja unter welchen Voraussetzungen beide Phänomene einer völkerrechtlichen Erfassung und Regelung zugänglich sind. In keinem Fall darf Cyberspace als „Wilder Westen“ oder gar „rechtliches Niemandsland“ angesehen werden.²¹ Vielmehr können die Regeln des analogen Völkerrechts grundsätzlich auch im digitalen Raum zur Anwendung gebracht werden.²² Welche Herausforderungen sich dabei stellen bzw. ob und wie die analogen Regelungen unter

20 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 205; s.a. UN, A/RES/2625 (XXV) vom 24.10.1970.

21 Ausdrücklich gegen eine Einordnung als „völkerrechtliches Niemandsland“ auch W. Heintschel von Heinegg, *Cyberspace – Ein völkerrechtliches Niemandsland?*, in: R. Schmidt-Radefeldt/C. Meissler (Hrsg.), *Automatisierung und Digitalisierung des Krieges, Drohnenkrieg und Cyberwar als Herausforderung für Ethik, Völkerrecht und Sicherheitspolitik*, Baden-Baden 2012, S. 159 (159 f.).

22 Dazu A. v. Arnould, *Völkerrecht*, 4. Aufl., Heidelberg 2019, S. 389 ff. Rn. 859 ff., insb. S. 392 ff. Rn. 864 ff.; D. J. Svantesson, *A New Jurisprudential Framework for Jurisdiction*, *AJIL Unbound* 109 (2015), 69; S. Schmahl, *Zwischenstaatliche Kompetenzabgrenzung*

Umständen angepasst oder modifiziert werden müssen, soll im Folgenden untersucht werden.

I. Ausgangspunkt: Cyberspace und Völkerrecht

Was ist also Cyberspace und warum stellt er das Völkerrecht vor besondere Herausforderungen? Eine allgemeinverbindliche völkerrechtliche Definition von Cyberspace gibt es bislang nicht. Als Begriff taucht „Cyberspace“ zum ersten Mal in zwei Science-Fiction-Werken von William Gibson auf, die 1982 und 1984 erschienen sind.²³ Umschreiben lässt sich Cyberspace am ehesten als „der virtuelle Raum, der durch die entsprechenden Verbindungen zwischen Sender und Empfänger und deren Computer hergestellt wird“ oder aber als „kommunikativer Raum zwischen Computern und fließenden Netzwerken“.²⁴ Eine ganz ähnliche Definition findet sich im Glossar der Cybersicherheitsstrategie für Deutschland des Bundesministeriums des Inneren, für Bau und Heimat (BMI) aus dem Jahr 2021:

„Der Cyberraum ist der virtuelle Raum aller weltweit auf Datenebene vernetzten beziehungsweise vernetzbaren informationstechnischen Systeme. Dem Cyberraum liegt als öffentlich zugängliches Verbindungsnetz das Internet zugrunde, das durch beliebige andere Datennetze erweitert werden kann.“²⁵

Als virtueller und durch weltweite Vernetzung geprägter Raum zeichnet sich Cyberspace insbesondere durch seine Ubiquität, also seine Ungebundenheit an einen bestimmten Standort aus. Obwohl einzelne Aspekte des virtuellen Raums wie beispielsweise Standorte von Computern oder Servern lokalisierbar sind, liegt der eigentliche Kern des Cyberspace in seiner dezentralen Vernetzung. Während im Völkerrecht klassischerweise eine Anknüpfung an staatliche Territorien und dem damit verbundenen Regelungsanspruch des

im Cyberspace, AVR 47 (2009), 284 (291 ff.); s.a. R. Uerpmann-Wittzack, GLJ 11 (2010), 1245 (1258).

23 W. Gibson, *Burning Chrome*, New York 1982; W. Gibson, *Neuromancer*, New York 1984; zur Urheberschaft Gibsons siehe Hobe, *Völkerrecht* (Fn. 13), S. 428.

24 Hobe, *Völkerrecht* (Fn. 13), S. 428.

25 BMI, *Cybersicherheitsstrategie für Deutschland 2021*, S. 133, abrufbar unter: <https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2021/09/cybersicherheitsstrategie-2021.pdf> (zuletzt abgerufen am 23.02.2023).

jeweiligen Staates möglich ist,²⁶ können im Cyberspace Ursprung, Zwischenstationen und Effekt einer Handlung räumlich weit auseinanderfallen und dennoch innerhalb kürzester Zeit verschiedene Staaten betreffen.²⁷ Um es mit den Worten des Völkerrechtlers *Andreas von Arnould* zu sagen: „Der virtuelle Raum des Cyberspace und der geographische Raum des Staates überlappen sich, passen aber nicht aufeinander.“²⁸

Wie bereits gesagt, sollte Cyberspace aber nicht vorschnell als völkerrechtliches „Niemandland“ eingeordnet werden. Vielmehr geben die Staaten ihre grundsätzlich umfassenden Regelungsansprüche für das eigene Territorium auch im Angesicht von Cyberspace nicht auf.²⁹ In Bezug auf das völkerrechtliche Selbstverteidigungsrecht nennen beispielsweise die Vereinigten Staaten in der *National Cyber Strategy* von 2018 ausdrücklich auch militärische Mittel als mögliche Antwort auf Cyberattacken.³⁰ Auch NATO-Generalsekretär *Jens Stoltenberg* hat keinen Zweifel daran gelassen, dass die NATO den Cyberspace als Domäne der Kriegführung auffasst und bereit ist, bei schweren Cyberattacken auch Maßnahmen der kollektiven Selbstverteidigung zu ergreifen.³¹ Inwieweit diese erklärte Bereitschaft auch auf eine völkerrechtliche Grundlage gestützt werden kann, gilt es noch im Einzelnen zu klären. Zuvor muss aber noch einmal das Bewusstsein für die zentrale Herausforderung geschärft werden, mit der sich das Völkerrecht bei Cyberoperationen konfrontiert sieht.

26 O. Diggelmann/N. Hadorn, Gewalt- und Interventionsverbot bei Cyberangriffen, Ausgewählte Schlüsselfragen, in: C. Schubel/S. Kirste/P. C. Müller-Graff/M. Diggelmann/U. Hufeld (Hrsg.), Jahrbuch für Vergleichende Staats- und Rechtswissenschaften – 2016/2017, Baden-Baden, S. 255 (260); zu Fragen der Souveränität und Cyberspace siehe nur N. Tsagourias, legal status, in: N. Tsagourias/R. Buchan (Hrsg.), Research Handbook on International Law and Cyberspace, 2. Aufl., Cheltenham 2021, S. 9 ff.; allgemein zum Umgang mit Entterritorialisierungstendenzen siehe K. Schmalenbach, VVDStRL 76 (2017), 245 sowie J. Bast, VVDStRL 76 (2017), 277 (289 ff.).

27 Dazu nur v. Arnould, Völkerrecht (Fn. 22), S. 389 Rn. 859; allgemein zum Umgang mit Entterritorialisierungstendenzen im Unions- und Völkerrecht siehe K. Schmalenbach, VVDStRL 76 (2017), 245 sowie J. Bast, VVDStRL 76 (2017), 277 (289 ff.).

28 A. v. Arnould, Völkerrecht (Fn. 22), S. 389 Rn. 859.

29 S. Shackleford/A. Craig, Beyond the New 'Digital Divide': Analyzing the Evolving Role of National Governments in Internet Governance and Enhancing Cybersecurity, in: StanfordJIL 50 (2014), 114. Zu den möglichen Anknüpfungspunkten siehe die Nachweise in Fn. 22.

30 National Cyber Strategy of the United States of America, September 2018, abrufbar unter: <https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf> (zuletzt abgerufen: 23.02.2023).

31 J. Stoltenberg, Nato will defend itself, in: Prospect, Cyber resilience vom 27.08.2019, abrufbar unter: <https://www.prospectmagazine.co.uk/world/nato-will-defend-itself-summit-jens-stoltenberg-cyber-security> (zuletzt abgerufen: 23.02.2023).

Cyberspace stellt das herkömmliche territoriale oder raumbezogene Denken im Völkerrecht deshalb vor Probleme, weil er sich aufgrund seiner Ubiquität der Zuordnung zu einer raumbezogenen staatlichen Jurisdiktionsgewalt jedenfalls in Teilen zu entziehen scheint. So banal diese Feststellung klingt: Die eigentliche Schwierigkeit im Umgang mit Cyberoperationen in völkerrechtlicher Hinsicht liegt in der fehlenden räumlichen Zuordnung. Dies lässt sich anhand eines einfachen Vergleichs von konventionellen und virtuellen Angriffen verdeutlichen. Konventionelle Militärschläge eines Staates A auf das Territorium eines Staates B lassen sich regelmäßig ohne Weiteres völkerrechtlich einordnen. Gleiches gilt für die Frage des Selbstverteidigungsrechts. Die Tatsache, dass moderne Kriegführung über Telekommunikations- oder Zielführungssysteme in weiten Teilen computergestützt abläuft, macht einen konventionellen Militärschlag noch nicht zu einer für die völkerrechtliche Einordnung problematischen Cyberoperation. Gleiches gilt für einen konventionellen Angriff auf Telekommunikations- oder sonstige IT-Einrichtungen. Die eigentliche Herausforderung für das Völkerrecht besteht in einem Angriff, der – ungeachtet seiner möglichen Auswirkungen auf die reale Welt – vollständig im oder durch den digitalen bzw. virtuellen Raum abläuft. Besonderes Augenmerk muss deshalb auf jenen Operationen liegen, „die Wirkungen *im* oder *durch* den Cyberraum zeitigen oder zu zeitigen bestimmt sind.“³² Wie bereits eingangs erwähnt, ist die Bandbreite möglicher Angriffe auf bzw. durch Computerbetriebssysteme mittels Viren, Würmern, Trojanischen Pferden, Logischen Bomben, Backdoor oder Denial of Service-Attacken enorm.³³ Eine ganz andere Frage ist es aber, ob derartige „Hacker-Angriffe“ auch in völkerrechtlicher Hinsicht als Angriff einzuordnen sind, der eine entsprechende staatliche Reaktion im Wege einer militärischen Selbstverteidigung rechtfertigen würde.

32 W. Heintschel von Heinegg, Völkerrechtliche Fragen im Cyber- und Informationsraum, Verteidigungsausschuss des Deutschen Bundestags, Öffentliche Anhörung vom 14. Dezember 2018, Ausschussdrucksache 19(12)941 vom 08.12.2020 – 19/3494 5410, S. 1 ff.; zum Begriff des Cyber-Angriffs s.a. R. Nguyen, Navigating Jus Ad Bellum in the Age of Cyber Warfare, CALIF. L. REV. 101 (2013), 1079 (1085 ff., insb. 1088).

33 Zu den verschiedenen Angriffsformen siehe die Nachweise in Fn. 8.

II. Möglichkeiten und Grenzen des Selbstverteidigungsrechts bei Cyberattacken

Cyberattacken eröffnen nur dann das Recht auf Selbstverteidigung, wenn im jeweiligen Einzelfall die Voraussetzungen von Art. 51 UN-Charta erfüllt sind. Ausweislich des klaren Wortlauts setzt das Selbstverteidigungsrecht zunächst das Vorliegen eines „bewaffneten Angriffs gegen einen Mitgliedstaat der Vereinten Nationen“ voraus (1.). Dieser Angriff – und das ist gerade bei Cyberattacken ein großes Problem – muss von einem Staat ausgehen oder diesem jedenfalls zugerechnet werden können (2.). Sollten beide Voraussetzungen erfüllt sein, müssen schließlich noch die Grenzen des Selbstverteidigungsrechts in quantitativer und zeitlicher Hinsicht gewahrt werden (3.).

1. Der „bewaffnete Angriff“ als Ausgangspunkt

Ein „bewaffneter Angriff“ bildet den Ausgangspunkt jeder staatlichen Selbstverteidigung. Es versteht sich zunächst von selbst, dass eine Selbstverteidigung von vornherein ausscheidet, wenn der vermeintliche „Angreifer“ selbst im Wege der Selbstverteidigung oder anderweitig gerechtfertigt handelt. Genauer wäre es daher von einem *rechtswidrigen* bewaffneten Angriff als Grundvoraussetzung zu sprechen.³⁴ Allerdings liegt die völkerrechtliche Problematik weniger in der Einordnung der Rechtmäßigkeit als vielmehr in der Einordnung einer Handlung als „bewaffneter Angriff“. Obwohl es die zentrale Voraussetzung ist, bleibt der Begriff des „bewaffneten Angriffs“ oder der „armed attack“ nämlich nach wie vor durch eine gewisse Unschärfe geprägt.³⁵ Eine wichtige Orientierung für die Praxis bietet die Aggressionsdefinition der Generalversammlung der Vereinten Nationen vom 14. Dezember 1974.³⁶ In Art. 3 dieser Resolution werden verschiedene Handlungen beispielhaft als Angriffshandlung definiert. Darunter fallen neben der Invasion oder dem Angriff der Streitkräfte eines Staates auf das Hoheitsgebiet eines anderen Staates unter anderem auch die Beschießung oder Bombardierung fremden Hoheitsgebietes, der Einsatz von Waffen jeder Art durch einen Staat gegen das Hoheitsgebiet eines anderen Staates, die Blockade von Häfen oder Küsten eines anderen Staates sowie der Angriff der Streitkräfte eines Staates auf die Land-, See- oder Luftstreitkräfte oder auf die See- und Luftflotte eines anderen

34 v. *Arnauld*, Völkerrecht (Fn. 22), S. 500 Rn. 1090.

35 *Hobe*, Völkerrecht (Fn. 13), S. 212.

36 UN Generalversammlung, A/RES/3314 (XXIX) vom 14.12.1974.

Staates. Obwohl diese nach Art. 4 der Resolution nicht abschließende Liste von Handlungen eine gute erste Orientierung bieten kann, sollte man bedenken, dass sich der Aggressionsbegriff dieser Resolution nur auf die Kompetenz des UN-Sicherheitsrats nach Art. 39 UN-Charta bezieht und ausdrücklich nicht als Definition für den „bewaffneten Angriff“ im Sinne des Art. 51 UN-Charta gedacht war.³⁷ Es besteht allerdings weitgehende Einigkeit, dass der Aggressionsbegriff zur Bestimmung eines Kernbereichs des „bewaffneten Angriffs“ herangezogen und der „bewaffnete Angriff“ im Übrigen durch Rechtsprechung und Praxis konkretisiert werden kann.³⁸ Im Allgemeinen scheint heute Folgendes anerkannt: Zunächst stellt nicht jede Gewaltanwendung einen „bewaffneten Angriff“ dar, sondern nur eine solche, die als massiver, koordinierter Militärschlag gegen einen anderen Staat eine gewisse Intensität erreicht.³⁹ So hat auch der Internationale Gerichtshof im *Nicaragua*-Fall bloße Grenzscharmützel zwar als Verstoß gegen das Gewaltverbot, nicht aber als einen das Selbstverteidigungsrecht auslösenden „bewaffneten Angriff“ eingeordnet.⁴⁰ Ob ein „bewaffneter Angriff“ vorliege, sei anhand des jeweiligen Einzelfalls zu bewerten, wobei es sowohl auf das Ausmaß bzw. die Größenordnung des Anschlages („scale“) als auch auf seine (Aus-)Wirkungen („effects“) ankomme.⁴¹ Wann Ausmaß und Wirkung die Schwelle des „bewaffneten Angriffs“ überschreiten, ist naturgemäß eine im Einzelfall schwierig zu beurteilende Frage. So hat beispielsweise der UN-Sicherheitsrat das Ausmaß wiederholter sogenannter „gestreuter Bagatellangriffe“ der PLO gegenüber Israel nicht als ausreichend angesehen.⁴² Umgekehrt hat der Inter-

37 A. Randelzhofer/G. Nolte, in: Simma/Khan/Nolte/Paulus (Fn. 14), Art. 51 Rn. 17; dazu und zum Folgenden auch v. Arnould, *Völkerrecht* (Fn. 22), S. 498 Rn. 1085.

38 Hobe, *Völkerrecht* (Fn. 13), S. 212; T. Stein/C. v. Buttlar/M. Kotzur, *Völkerrecht*, 14. Aufl., München 2017, S. 292 Rn. 784.

39 Siehe nur Herdegen, *Völkerrecht* (Fn. 13), S. 281 Rn. 22; Hobe, *Völkerrecht* (Fn. 13), S. 212 f.; Stein/v. Buttlar/Kotzur, *Völkerrecht* (Fn. 38), S. 292 Rn. 784; teilweise anders aber W. Heintschel von Heinegg, in: K. Ipsen, *Völkerrecht*, 7. Aufl., München 2018, § 56 Rn. 6 ff.

40 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 195.

41 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, §§ 191, 195; IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161, §§ 51 u. 64. Ausf. auch T. Ruys, „Armed Attack“ and Article 51 of the UN Charter: Evolutions in Customary Law and Practice, Cambridge 2010, passim.

42 UN Sicherheitsrat, S/RES/490(1981) vom 21.07.1981; S/RES/501(1982) vom 25.02.1982; S/RES/509 vom 06.06.1982; dazu Stein/v. Buttlar/Kotzur, *Völkerrecht* (Fn. 38), S. 292 Rn. 786.

nationale Gerichtshof eine kumulative Gesamtbetrachtung nicht vollständig ausgeschlossen.⁴³ Generell dürfte bei hohen Opferzahlen die Schwelle zum „bewaffneten Angriff“ überschritten sein. Gleichzeitig wird gerade im Hinblick auf die „Auswirkungen“ der fraglichen Handlung oftmals eine hohe Hürde angesetzt, die auf eine Bedrohung der politischen Unabhängigkeit oder Souveränität des betroffenen Staates abstellt.⁴⁴ Eine derart hohe Hürde mag auf den ersten Blick unbefriedigend erscheinen. Obwohl ein klarer aggressiver Akt vorliegt, der möglicherweise auch gegen das Gewaltverbot verstößt, wird dem „Opfer“ der Aggression das Recht zur Selbstverteidigung verwehrt. Diesem Umstand versuchen einige Stimmen zu begegnen, indem sie das Selbstverteidigungsrecht nicht vollständig ausschließen, sondern „nur“ in der Rechtsfolge beschränken wollen. Wird nach dieser Ansicht die Schwelle des „bewaffneten Angriffs“ nach Ausmaß und Wirkung nicht überschritten, dürfe der betroffene Staat auf eine „*forcible countermeasure*“ unter strenger Wahrung der Verhältnismäßigkeit zurückgreifen.⁴⁵ Die Etablierung eines solchen „kleinen Selbstverteidigungsrechts“⁴⁶ ist mit Blick auf eine mögliche Eskalation des Konflikts aber mit Vorsicht zu genießen.⁴⁷ Wie auch der Internationale Gerichtshof im *Nicaragua*-Fall noch einmal betont hat, ist es das Ziel der UN-Charta, gewalttätige Auseinandersetzungen zwischen den Staaten der Vereinten Nationen im größtmöglichen Umfang auszuschließen.⁴⁸ Im Falle des Nichterreichens der Schwelle des „bewaffneten Angriffs“ ist der betroffene Staat somit mangels Ausnahme vom Gewaltverbot auf friedliche Gegenmaßnahmen beschränkt.⁴⁹ In Betracht kommt neben Maßnahmen im Bereich des Diplomaten- bzw. Konsularrechts insbesondere die

43 Vgl. IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161 § 64.

44 Dazu *Stein/v. Buttlar/Kotzur*, Völkerrecht (Fn. 38), S. 292 f. Rn. 787.

45 Sondervotum B. Simma zum Urteil vom 06.11.2003, IGH, *Oil Platforms* (Iran v. USA), ICJ-Reports 2003, 161, 324 §§ 12 ff.; dazu auch *M. Kowalski*, Original Sin Reaffirmed, *PolyIL* 36 (2016), 37 (43 ff.).

46 Dazu *A. Verdross/B. Simma*, Universelles Völkerrecht, Theorie und Praxis, 3. Aufl., Berlin 1984, S. 289 f. § 473; *Herdegen*, Völkerrecht (Fn. 13), S. 273 ff. Rn. 25.

47 *v. Arnould*, Völkerrecht (Fn. 22), S. 499 Rn. 1087; kritisch auch *A. Randelzhofer/G. Nolte*, in: *Simma/Khan/Nolte/Paulus* (Fn. 14), Art. 51 Rn. 8.

48 *Stein/v. Buttlar/Kotzur*, Völkerrecht (Fn. 38), S. 294 Rn. 790 unter Hinweis auf IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 127, § 14.

49 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, §§ 191, 195; IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161, § 51.

Verhängung von Wirtschaftssanktionen.⁵⁰ Eine Klassifizierung als „bewaffneter Angriff“ ist also schon bei konventionellen bzw. analogen Angriffen nicht unproblematisch. Noch einmal komplizierter wird es dann, wenn Cyberattacken rechtlich eingeordnet werden sollen.

Können Cyberattacken einen „bewaffneten Angriff“ im Sinne des Art. 51 UN-Charta darstellen? Zur Beantwortung sei zunächst auf ein Gutachten des Internationalen Gerichtshofs verwiesen, wonach die Ausübung des Rechts auf Selbstverteidigung nicht von der Art der Waffe abhängt, mit der ein Angriff verübt wurde.⁵¹ Dies entspricht auch der Staatenpraxis, wonach auch chemische, biologische oder atomare Attacken als „bewaffnete Angriffe“ qualifiziert werden können.⁵² Eine andere Frage ist hingegen, ob der Wortlaut von Art. 51 UN-Charta ein physisches Element in Bezug auf die Waffe fordert oder ob auch rein elektronische oder virtuelle Waffen erfasst sind.⁵³ Die Formulierung „bewaffnet“ bzw. „armed“ spricht eher gegen eine unmittelbare Erfassung nicht-physischer Waffen. Dies darf aber nicht zu einem Ausschluss des Selbstverteidigungsrechts führen, da die Anwendung des naturgegebenen Selbstverteidigungsrechts nicht davon abhängen kann, wie ein möglicher Angriff verursacht bzw. ein entstandener Schaden vermittelt wurde. Dementsprechend kann auch ein rein virtueller Angriff grundsätzlich zu einer entsprechenden bzw. analogen Anwendung von Art. 51 UN-Charta führen.⁵⁴

Damit ist aber noch lange nicht geklärt, unter welchen Voraussetzungen eine Cyberattacke einen „bewaffneten Angriff“ darstellt, der eine Selbstverteidigung rechtfertigen würde. Entsprechend der allgemeinen Annäherung an den Begriff kommt es nach überwiegender Auffassung auch bei Cyberattacken auf Ausmaß und Wirkung an.⁵⁵ Damit eine Cyberattacke die Schwelle

50 v. *Arnauld*, Völkerrecht (Fn. 22), S. 499 Rn. 1087 u. S. 256 f. Rn. 421; s.a. *H. Krieger*, Krieg gegen anonymous, AvR 50 (2010), 1 (14 ff.); speziell zu Wirtschaftssanktionen s.a. *G. Hafner*, Völkerrechtliche Grenzen und Wirksamkeit von Sanktionen gegen Völkerrechtssubjekte, ZaöRV 76 (2016), 391.

51 IGH, *Legality of the Threat of Use of Nuclear Weapons* (Advisory Opinion), Gutachten vom 08.07.1996, ICJ-Reports 1996, 226, §§ 38 f.

52 Dazu *M. N. Schmitt* (Hrsg.), *Tallinn Manual 2.0 on the international law applicable to cyber operations*, Cambridge 2017, S. 340 Rule 71 Erläuterung 4 m.w.N.

53 Für eine Erfassung v. *Arnauld*, Völkerrecht (Fn. 22), S. 498 f. Rn. 1086; ausf. auch *A. Randelzhofer/G. Nolte*, in: *Simma/Khan/Nolte/Paulus* (Fn. 14), Art. 51 Rn. 43 m.w.N.

54 *Hobe*, Völkerrecht (Fn. 13), S. 248; zum Argument der Wortherkunft s.a. *Roscini*, *Cyber operations* (Fn. 13), S. 297 (301 ff.).

55 Einen Überblick über andere Kriterien bieten *L. A. Petersen*, *Cyberangriffe - Definition, Regulierung, Pönalisierung*, Göttinger Rechtszeitschrift 1/2020, 25 (28 ff.) sowie *S. Pan-*

eines „bewaffneten Angriffs“ im Sinne des Art. 51 UN-Charta (analog) überschreitet, müssen ihre physischen Auswirkungen, das heißt insbesondere die hervorgerufenen Schäden, mit denen eines konventionellen „bewaffneten Angriffs“ vergleichbar sein.⁵⁶ Insoweit werden also in erster Linie die Auswirkungen der virtuellen Attacke auf die reale Welt in den Blick genommen und daraufhin geprüft, ob sie in Ausmaß und Wirkung einem konventionellen Angriff entsprechen. Das Tallinn Manual, ein internationales Handbuch für rechtlich zulässige Methoden der Kriegführung, welche nun auch Regeln über die Anwendung des Völkerrechts auf Cyberoperationen zum Inhalt hat,⁵⁷ bietet für diesen Vergleich eine nicht abschließende Liste von Kriterien an, die neben dem Grad des Schadens auch einige qualitative Elemente der Cyberoperation in den Blick nimmt. Im Einzelnen nennt das Manual *severity*, *immediacy*, *directness*, *invasiveness*, *measurability of effects*, *military character*, *state involvement* und *presumptive legality*.⁵⁸ Neben dieser Liste, die das Tallinn Manual in erster Linie nur auf das Gewaltverbot nach Art. 2 Ziff. 4 UN-Charta bezieht, bietet es auch im Zusammenhang mit dem Selbstverteidigungsrecht und der dort relevanten Einordnung als „bewaffneter Angriff“ einige Hilfestellung.⁵⁹ In Anlehnung an die auch bei konventionellen Angriffen herangezogene Aggressionsdefinition können Cyberoperationen gegen die militärische Infrastruktur eines Staates, welche die Verteidigungsbereitschaft erheblich einschränken, als bewaffneter Angriff eingeordnet werden. Ein einfaches Beispiel wäre ein Cyberangriff gegen die IT eines Kriegsschiffes oder Kampfflugzeugs als Vorbereitung für einen konventionellen Angriff.⁶⁰ Ist allerdings die Verteidigungsbereitschaft nicht unmittelbar be-

grazzi, Self-defence against Cyberattack?, Digital and kinetic defence in light of article 51 UN-Charter, Policy Brief, ICT for Peace Foundation, Genf 2021, S. 12 ff., abrufbar unter: <https://ict4peace.org/wp-content/uploads/2021/03/ICT4Peace-2021-Cyberattacks-and-Article51-1.pdf> (zuletzt abgerufen: 23.02.2023).

56 Siehe nur *v. Arnould*, Völkerrecht (Fn. 22), S. 498 f. Rn. 1086; *Hobe*, Völkerrecht (Fn. 13), S. 248; *A. Randelzhofer/G. Nolte*, in: Simma/Khan/Nolte/Paulus (Fn. 14), Art. 51 Rn. 42 f. m.w.N. Ausf. auch *Focarelli*, Self-defence (Fn. 1), S. 317 (326 ff.) sowie *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 339 ff. Rule 71 ff.

57 *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 1 ff.

58 *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 333 ff. Rule 69 Erläuterung 9 f.

59 *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 339 ff. Rule 71; siehe zum Zusammenhang von Gewaltverbot und bewaffneten Angriff auch *Focarelli*, Self-defence (Fn. 1), S. 317 (327 f.).

60 *Herdegen*, Völkerrecht (Fn. 13), S. 283 Rn. 24; zu möglichen Formen und Abstufungen s.a. *Focarelli*, Self-defence (Fn. 1), S. 317 (329 ff.) sowie *F. M. E. Oorsprong/P. A. L. Duchaine/B. M. J. Pijpers*, Armed attack in Cyberspace, Clarifying and Assessing when Cyber-

troffen, wird auch die Schwelle zum „bewaffneten Angriff“ regelmäßig nicht überschritten sein. Dies gilt insbesondere für rein netzinterne Vorgänge, wie beispielsweise im Falle des bloßen Ausspähens von Daten oder der kurzfristigen Blockade von Regierungsseiten.⁶¹ Obwohl diese und ähnlich gelagerte Situationen also nicht zur Selbstverteidigung ermächtigen, können sie gleichwohl als Verstoß gegen das Interventionsverbot zu qualifizieren sein und dementsprechend friedliche Gegenmaßnahmen ermöglichen.⁶²

Anders sieht die Situation wiederum aus, wenn man Cyberattacken auf zivile Einrichtungen oder Objekte betrachtet. Denkbare Szenarien sind insbesondere Cyberoperationen, die kritische Infrastruktur ins Visier nehmen. Für Deutschland nennt § 2 Abs. 2 des Gesetzes über das Bundesamt für Sicherheit in der Informationstechnik (BSI-Gesetz), auf welches auch die Cybersicherheitsstrategie des BMI Bezug nimmt,⁶³ als kritische Infrastrukturen:

Einrichtungen, Anlagen oder Teile davon,

1. die den Sektoren Energie, Informationstechnik und Telekommunikation, Transport und Verkehr, Gesundheit, Wasser, Ernährung, Finanz- und Versicherungswesen sowie Siedlungsabfallentsorgung angehören und
2. von hoher Bedeutung für das Funktionieren des Gemeinwesens sind, weil durch ihren Ausfall oder ihre Beeinträchtigung erhebliche Versorgungsengpässe oder Gefährdungen für die öffentliche Sicherheit eintreten würden.

Cyberoperationen gegen derartige Einrichtungen, die zu nicht unerheblichen Verletzungen, Todesfällen oder substanziellen Sachschäden oder Zerstörungen

Attacks Trigger the Netherlands' Right of Self-Defence, Amsterdam Law School Legal Studies Research Paper No. 2021–29, S. 16 ff.

61 v. *Arnauld*, Völkerrecht (Fn. 22), S. 498 f. Rn. 1086 sowie *Hobe*, Völkerrecht (Fn. 13), S. 248, jeweils m.w.N. Ausf. zur völkerrechtlichen Einordnung von Cyber-Spionage *R. Buchan/I. Navarette*, Cyber espionage and international law, in: N. Tsagourias/R. Buchan (Hrsg.), *Research Handbook on International Law and Cyberspace*, 2. Aufl., Cheltenham 2021, S. 231 ff.

62 Ausf. *Schmitt* (Hrsg.), *Tallinn Manual 2.0* (Fn. 52), S. 168 ff. Rule 32 sowie S. 312 ff. Rule 66; s.a. *K. Ziolkowski*, *Peacetime Cyber Espionage – New Tendencies in Public International Law*, in: K. Ziolkowski (Hrsg.), *Peacetime Regime for State Activities in Cyberspace*, *International Law, International Relations and Diplomacy*, Tallinn 2013, S. 425 (457 ff.) sowie *T. D. Gill*, *Non-Intervention in the Cyber-Context*, in: K. Ziolkowski (Hrsg.), *Peacetime Regime for State Activities in Cyberspace*, *International Law, International Relations and Diplomacy*, Tallinn 2013, S. 217 ff.

63 BMI, *Cybersicherheitsstrategie für Deutschland 2021*, S. 15 f., 52 u. 81, abrufbar unter: <https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2021/09/cybersicherheitsstrategie-2021.pdf> (zuletzt abgerufen am 23.03.2023).

gen führen, können ebenfalls als „bewaffneter Angriff“ im Sinne des Art. 51 UN-Charta eingeordnet werden.⁶⁴ Fehlt es an derartigen klaren oder unmittelbaren Folgen, kann die Abgrenzung schwierig sein. Beispielsweise hat ein Cyberangriff auf internationale Finanzmärkte zweifelsfrei enorme wirtschaftliche Folgen, führt aber nicht zwangsläufig zu unmittelbaren physischen Schäden. Letztlich kommt es auf eine einzelfallbezogene Betrachtung an, bei der grundsätzlich alle vorhersehbaren Effekte sowohl im unmittelbar betroffenen Staat als auch in Drittstaaten berücksichtigt werden müssen.⁶⁵ Umstritten bleibt dabei, inwieweit die Intention des Angreifers zu berücksichtigen ist.⁶⁶ Letztlich bleiben aber Ausmaß und Wirkung der Cyberoperation und ihre Vergleichbarkeit zu einem konventionellen Angriff entscheidend. Dabei ist auch in Rechnung zu stellen, dass in der globalisierten Welt Cyberattaken auf die kritische Infrastruktur eines Landes schnell überregionale oder sogar weltweite Auswirkungen haben können. Umgekehrt sollte mit Blick auf die hohe Bedeutung des Gewaltverbots nicht vorschnell ein „bewaffneter Angriff“ angenommen werden.

Als erstes Zwischenergebnis lässt sich aber somit festhalten, dass Cyberattaken, sofern sie nach Ausmaß und Wirkung mit einem konventionellen Angriff vergleichbar sind, einen „bewaffneten Angriff“ im Sinne des Art. 51 UN-Charta darstellen können. Damit ist aber nur die Grundvoraussetzung des Selbstverteidigungsrechts erfüllt. Die eigentliche Herausforderung in faktischer und völkerrechtlicher Hinsicht liegt darin, zu beurteilen, von wem der Angriff ausgegangen ist und gegen wen sich dementsprechend auch eine Gegenmaßnahme zu richten hat.

2. Die eigentliche Herausforderung: Rückverfolgbarkeit und Zurechnung

Wie oben dargestellt, kann schon die Qualifizierung einer Cyberattacke als „bewaffneter Angriff“ eine im Einzelfall schwierig zu beantwortende

64 *Focarelli*, Self-defence (Fn. 1), S. 317 (330 ff.) sowie *A. Randelzhofer/G. Nolte*, in: Simma/Khan/Nolte/Paulus (Fn. 14), Art. 51 Rn. 43; s.a. *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 341 ff. Rule 71, Erläuterungen 8 und 12 ff.

65 Ausf. *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 342 ff. Rule 71, Erläuterung 12 ff. sowie *Dornbusch*, Kampfführungsrecht (Fn. 1), S. 20, 24 f., 81 ff., 102 ff. und 228 f.

66 In IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161, § 64 spricht der IGH von „the specific intention of harming“; dazu *S. Kolossa*, Der Cyberraum – Chance, Gefahrenraum und Waffe?, HuV 1 (2018), 151 (166) sowie *M. Roscini*, World Wide Warfare – Jus ad bellum and the Use of Cyber Force, in: A. v. Bogdandy/R. Wolfrum, Max Plank Yearbook of United Nations Law, Vol. 14 (2010), S. 85 (115 f.).

rechtliche Frage darstellen. Doch selbst wenn ein Staat unzweifelhaft von einer Cyberoperation betroffen ist und er sie aufgrund ihrer Auswirkungen zu Recht als „bewaffneten Angriff“ einordnet, steht der Staat vor dem nicht unerheblichen Problem, gegen wen er seine Gegenmaßnahmen richten soll. Die damit angesprochenen Fragen der Identifizierung des Angreifers sowie der Zurechnung zu einem Staat sind nicht nur in völkerrechtlicher, sondern auch und vor allem in technischer Hinsicht mit Schwierigkeiten verbunden.⁶⁷

So wird bei Cyberoperationen regelmäßig zunächst die technische Frage der schwierigen bis unmöglichen Rückverfolgbarkeit des Angriffs die eigentliche Herausforderung darstellen.⁶⁸ Die Funktionsweise des Cyberspace als weltweit verzweigtes Netzwerk lässt dem Angreifer zahllose technische Wege, entweder Ursprung und Route seines Angriffs komplett zu verbergen oder aber auch falsche Fährten zu legen. Selbst wenn eine Cyberattacke über Leitungen oder Computer eines bestimmten Staates verübt wurde, ist das zwar ein Indiz, aber noch kein vollständiger Nachweis darüber, ob dieser Staat tatsächlich der Ursprungsort des Angriffs war.⁶⁹ Eine große Rolle spielt auch der Faktor Zeit. Während die Attacke selbst in der Regel nur wenige Augenblicke dauern wird, können ihre Wirkungen unter Umständen erst mit Verzögerung eintreten oder aber erst später als Folge einer Cyberattacke erkennbar werden.⁷⁰ Ein letztes in der Regel kaum zu überbrückendes Problem ist schließlich die sogenannte „Mensch-Maschine-Gap“.⁷¹ Selbst wenn sich der Weg eines Angriffs vollständig und innerhalb angemessener Zeit zu einem bestimmten Server oder Computer zurückverfolgen lässt

67 Zum Unterschied von Identifizierung und Zurechnung sowie dem technisch-politischen Rahmen siehe nur *Roscini*, World Wide Warfare (Fn. 66), S. 85 (96 ff.) sowie *N. Tsagourias*, Cyber attacks, self-defence and the problem of attribution, *Journal of Conflict and Security Law* Vol. 17 (2012), 229 (230 u. 233 ff.).

68 Dazu ausf. *Schulze*, Cyber-„War“ (Fn. 6), S. 36 ff.; s.a. *Heintschel von Heinegg*, Cyberspace (Fn. 21), S. 159 (171); *Kolossa*, Cyberraum (Fn. 66), 167 sowie die Nachweise in Fn. 67.

69 *Kolossa*, Cyberraum (Fn. 66), 167; s.a. *Schmitt* (Hrsg.), Tallinn Manual on the International Law Applicable to Cyber warfare, Cambridge 2013, S. 34 ff. Rule 7 ff. sowie *S. Pangrazzi*, Self-defence against Cyberattacks?, Digital and kinetic defence in light of article 51 UN-Charter, Policy Brief, ICT for Peace Foundation, Genf 2021, S. 17 f., abrufbar unter: <https://ict4peace.org/wp-content/uploads/2021/03/ICT4Peace-2021-Cyberattacks-and-Article51-1.pdf> (zuletzt abgerufen: 23.02.2023).

70 *Schulze*, Cyber-„War“ (Fn. 6), S. 46 f.; zum Problem der Verzögerung auch *C. Schaller*, Internationale Sicherheit und Völkerrecht im Cyberspace, Für klare Regeln und mehr Verantwortung, SWP-Studie 2014/S 18 vom 23.10.2014, S. 21 f.

71 Dazu und zum Folgenden *Schulze*, Cyber-„War“ (Fn. 6), S. 47.

und somit der Ursprungsort feststeht, ist damit noch nicht geklärt, welcher Mensch an diesem Computer saß und in welcher Funktion er tätig wurde.⁷² Diese Frage wird sich in der Praxis regelmäßig auch nicht oder jedenfalls nicht innerhalb kurzer Zeit klären lassen. In technischer Hinsicht besteht damit eine nur schwer zu überwindende Einschränkung für das Selbstverteidigungsrecht. Kann der Angreifer nicht zweifelsfrei identifiziert werden, ist eine Verteidigung schon rein praktisch ausgeschlossen. Ob und vor allem wie dieser technischen Unzulänglichkeit mit neuen Ansätzen und Mitteln des Völkerrechts begegnet werden kann, wird im Rahmen des Ausblicks am Ende des Beitrags erläutert. Zunächst soll die Problematik der Rückverfolgbarkeit aber zurückgestellt werden und der Fokus auf den weiteren Voraussetzungen und Grenzen des Selbstverteidigungsrechts liegen.

Ungeachtet der schwierigen Rückverfolgbarkeit muss sich die Selbstverteidigung grundsätzlich gegen den Staat richten, der den Angriff verübt hat oder dem ein eventuelles Handeln privater Gruppen oder Akteure zugerechnet werden kann. Einen Sonderfall bildet die umstrittene Selbstverteidigung gegen private Akteure, die unabhängig von einem Staat handeln, namentlich terroristische Gruppen. Auch Cyberangriffe können logischerweise von Staaten oder privaten Akteuren ausgeführt werden, wobei Handlungen von nichtstaatlichen Akteuren sicherlich die größte Herausforderung darstellen. In Bezug auf konventionelle bzw. analoge „bewaffnete Angriffe“ hat der Internationale Gerichtshof im *Nicaragua*-Fall geäußert, dass die Entsendung bewaffneter Banden im Sinne des Art. 3 lit. g der Aggressionsdefinition einem Staat zugerechnet und somit zu einem Selbstverteidigungsrecht führen kann.⁷³ Für eine Zurechnung darf der Staat aber nicht nur als bloßer finanzieller oder logistischer Unterstützer erscheinen. Vielmehr muss er im Sinne einer Tatherrschaft so über die Handlungen der privaten Gruppen bestimmen, dass sie gewissermaßen als sein „verlängerter Arm“ agieren.⁷⁴ Besteht keine solche Verbindung, scheidet eine Selbstverteidigung unmittelbar gegen Angriffe von privaten Akteuren grundsätzlich aus. Dieses Verständnis hat sich im Nachgang zu den Terroranschlägen vom 11. September 2001 aber teilweise

72 *Heintschel von Heinegg*, Cyberspace (Fn. 21), S. 159 (172); *Kolossa*, Cyberraum (Fn. 66), 167.

73 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 195.

74 *v. Arnould*, Völkerrecht (Fn. 22), S. 499 f. Rn. 1088 f.; s.a. *Herdegen*, Völkerrecht (Fn. 13), S. 284 Rn. 25 f. sowie *Stein/v. Buttlar/Kotzur*, Völkerrecht (Fn. 38), S. 293 Rn. 788 f.

gewandelt.⁷⁵ Während der Internationale Gerichtshof wohl weiter an der herkömmlichen Sichtweise festhält,⁷⁶ gehen der UN-Sicherheitsrat, die Staatengemeinschaft und wohl auch der überwiegende Teil der Literatur davon aus, dass auch private Gewaltanwendung einen „bewaffneten Angriff“ darstellen kann, sofern sie das Ausmaß einer staatlichen Militäraktion erreicht. In diesem Fall würde sich die Selbstverteidigung gegen den Staat richten, der den Angreifern einen Unterschlupf – man spricht auch von einem „safe haven“ – gewährt hat.⁷⁷

Obwohl gerade diese letzte Ansicht nicht unumstritten ist, können die grundlegenden Gedanken der Zurechnung auch zur rechtlichen Einordnung von Cyberattacken herangezogen werden. Wenn eine Gruppe oder einzelne Personen einen Cyberangriff als verlängerter Arm eines Staates ausführen, ist auch insoweit eine Zurechnung möglich. Bei Cyberangriffen, die Private vom Territorium eines Staates verüben, ohne dass dieser den Angriff ausdrücklich autorisiert hat oder ihn jedenfalls duldet, bleibt die Lage umstritten.⁷⁸ Problematisch ist vor allem, wann dem Staat, von dessen Territorium ein Angriff verübt wird, in rechtlicher Hinsicht ein Vorwurf gemacht werden kann, der auch ein Selbstverteidigungsrecht des angegriffenen Staates rechtfertigen würde. Da dies auch mit der bereits problematisierten technischen Rückverfolgbarkeit zusammenhängt, werden beide Aspekte noch einmal am Ende des Beitrags im Zusammenhang mit den künftigen Entwicklungslinien des Völkerrechts aufgegriffen. Zunächst kann aber als zweites Zwischenergebnis festgehalten werden, dass eine Selbstverteidigung auch bei privaten Cyberattacken nicht von vornherein ausgeschlossen ist.

3. Grenzen des Selbstverteidigungsrechts

Wenn Cyberattacken also ungeachtet der schwierigen Rückverfolgbarkeit ein staatliches Selbstverteidigungsrecht legitimieren können, bleibt noch zu klä-

75 Hier und im Folgenden *Kolossa*, Cyberraum (Fn. 66), 166 und *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 344 Rule 71, Erläuterung 18, jeweils m.w.N.

76 IGH, *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory* (Advisory Opinion) Gutachten vom 09.07.2001, ICJ-Reports 2004, 136, § 139; IGH, *Armed Activities on the Territory of the Congo*, Merits, Urteil vom 19.12.2005, ICJ-Reports 2005, 168, §§ 146 f. u. 160.

77 *Hobe*, Völkerrecht (Fn. 13), S. 213; ausf. *Stein/v. Buttlar/Kotzur*, Völkerrecht (Fn. 38), S. 316 ff. Rn. 841 ff. u. 298 ff. Rn. 800.; s.a. die Nachweise in Fn. 75; kritischer *v. Arnould*, Völkerrecht (Fn. 22), S. 271 ff. Rn. 1120 ff.

78 Zu der Übertragbarkeit der Gedanken und dem Meinungsstand siehe nur *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 344 u. 347 Rule 71, Erläuterungen 16 ff. u. 23 ff.

ren, wie die konkrete Verteidigungshandlung ablaufen muss. Dies führt zu der abschließenden Frage, welchen Grenzen das staatliche Selbstverteidigungsrecht aus Art. 51 UN-Charta unterliegt.

Entsprechend seiner Zielsetzung, einen „bewaffneten Angriff“ auf das eigene Territorium *abzuwehren*, eröffnet das Selbstverteidigungsrecht nicht die Möglichkeit für eine Vergeltung. Es richtet sich alleine gegen einen aktuell andauernden bzw. *gegenwärtigen* Angriff. Eine Selbstverteidigung ist daher grundsätzlich nur möglich, wenn der Angriff bereits begonnen hat und noch nicht abgeschlossen ist. Dieses Gegenwärtigkeits- oder auch Unmittelbarkeitskriterium kann sowohl in Bezug auf den Beginn als auch das Ende des Angriffs bzw. der Selbstverteidigung problematisch sein. Zunächst ist das Kriterium nicht im Sinne einer strengen „Unverzüglichkeit“ zu verstehen. So geht die Gegenwärtigkeit regelmäßig nicht verloren, wenn ein Staat sich erst mit seinen Verbündeten abstimmt oder andere notwendige Vorbereitungs-handlungen für einen Gegenschlag trifft.⁷⁹ Mit anderen Worten: Es besteht hinsichtlich des Endes der Gegenwärtigkeit ein gewisser Spielraum. Umstritten ist, inwieweit dies auch beim Angriffsbeginn der Fall ist. Dies betrifft insbesondere die sogenannte „antizipatorische Selbstverteidigung“, also eine Verteidigung gegen einen unmittelbar bevorstehenden Angriff. Regelmäßig dürfte es einem Staat nicht zuzumuten sein, so lange abzuwarten, bis der Angriff tatsächlich begonnen hat. Um die damit verbundene Missbrauchsgefahr zu begrenzen, wird zur Bestimmung der Frage, wann die Unmittelbarkeit noch gewahrt ist, auf die sogenannte *Caroline-* oder *Webster-*Formel abgestellt. Demnach muss der Staat, der sich auf die Selbstverteidigung beruft, nachweisen, dass der Angriff unmittelbar bevorstand, überwältigend war und dass keine Wahl der Mittel und keine Zeit für weitere Beratungen blieb.⁸⁰ Die Beurteilung, wann diese Kriterien erfüllt sind, hängt naturgemäß vom jeweiligen Einzelfall ab. Letztlich spricht einiges dafür, dem angegriffenen Staat auch hier einen – begrenzten – Beurteilungsspielraum zuzugestehen. Andernfalls könnte der Angreifer allein die Regeln des Spiels bestimmen.

Das Erfordernis der Gegenwärtigkeit kann bei Cyberangriffen eine nicht unerhebliche Hürde darstellen. Wie bereits geschildert, spielt der Zeitfaktor eine wichtige Rolle bei Cyberattacken. Einerseits läuft die Angriffshandlung sehr schnell ab. Ihre Folgen können aber unter Umständen erst viel später

79 v. *Arnauld*, Völkerrecht (Fn. 22), S. 500 f. Rn. 1091 f.; *Hobe*, Völkerrecht (Fn. 13), S. 213; *Stein/v. Buttlar/Kotzur*, Völkerrecht (Fn. 38), S. 294 f. Rn. 792 f.

80 Dazu und zum Folgenden v. *Arnauld*, Völkerrecht (Fn. 22), S. 501 Rn. 1093 sowie *Hobe*, Völkerrecht (Fn. 13), S. 213 f., jeweils m.w.N.

Eintreten oder erkannt werden. Auch kann es eine lange Zeit dauern, bis der Angriff zurückverfolgt und so das Ziel der Selbstverteidigung identifiziert ist. Die soeben geschilderten Gedanken, also insbesondere die Möglichkeit einer antizipatorischen Selbstverteidigung und die Notwendigkeit, dem angegriffenen Staat eine gewisse Vorbereitungs- und Nachforschungszeit einzuräumen, dürften grundsätzlich auch bei Cyberoperationen anwendbar sein, wobei die Einzelheiten weiter umstritten sind.⁸¹ Auch bei Cyberoperationen ist zur Wahrung der Unmittelbarkeit aber entscheidend, dass der zeitliche und inhaltliche Bezug von Angriff und Verteidigungshandlung nicht vollständig verloren geht.⁸² In Anbetracht der schwierigen Rückverfolgbarkeit macht dies möglicherweise eine weitergehende Flexibilisierung des Unmittelbarkeitskriteriums erforderlich. Andererseits darf wiederum die drohende Missbrauchsgefahr nicht unterschätzt werden.

Als weitere Grenze des Selbstverteidigungsrechts muss der Grundsatz der Verhältnismäßigkeit beachtet werden. Obwohl diese Voraussetzung sich nicht aus dem Wortlaut des Art. 51 UN-Charta ergibt, hat der Internationale Gerichtshof sie mit Blick auf die gewohnheitsrechtlichen Grundlagen des Selbstverteidigungsrechts ausdrücklich auch für Art. 51 UN-Charta bestätigt.⁸³ Eine Verteidigungshandlung muss demnach erforderlich sein, es darf also kein milderes, gleich wirksames Mittel geben. Zudem darf die ergriffene Maßnahme nicht außer Verhältnis zu Umfang und Auswirkung des Angriffs stehen, sie muss also angemessen sein.⁸⁴ Beide Fragen sind naturgemäß nicht allgemein, sondern nur anhand eines konkreten Einzelfalls zu beantworten. Dabei dürfen nur rechtliche, nicht aber politische Kriterien eine Rolle spielen. Wichtige Gesichtspunkte sind u.a. die Art und Intensität des Angriffs sowie der Bewaffnung, das Ausmaß erwarteter Schäden auf beiden Seiten und die möglichen Auswirkungen auf Drittstaaten. Eine strenge Symmetrie zwischen

81 Zum Meinungsstand *Schmitt* (Hrsg.), *Tallinn Manual 2.0* (Fn. 52), S. 350 ff. Rule 73; s.a. *Focarelli*, *Self-defence* (Fn. 1), S. 317 (332 ff.).

82 Vgl. *Stein/v. Buttlar/Kotzur*, *Völkerrecht* (Fn. 38), S. 295 Rn. 793.

83 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 194; IGH, *Legality of the Threat of Use of Nuclear Weapons* (Advisory Opinion), Gutachten vom 08.07.1996, ICJ-Reports 1996, 226, § 41; IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161, §§ 43, 73 f. u. 76.

84 *A. Randelzhofer/G. Nolte*, in: *Simma/Khan/Nolte/Paulus* (Fn. 14), Art. 51 Rn. 57 f.; dazu und zum Folgenden auch *Stein/v. Buttlar/Kotzur*, *Völkerrecht* (Fn. 38), S. 295 ff. Rn. 794 ff.

Angriff und Verteidigung gibt es dabei nicht. Der verteidigende Staat darf auf die effektive und endgültige Beendigung des Angriffs zielen.⁸⁵

Gerade der letztgenannte Aspekt ist für Cyberoperationen besonders relevant. So kann ein angegriffener Staat auf einen Cyberangriff durchaus mit einem konventionellen Gegenschlag reagieren. Umgekehrt ist auch ein Cyber-Gegenschlag als Reaktion auf einen konventionellen Schlag denkbar.⁸⁶ Wiederum sind allein die rechtlichen Kriterien der Erforderlichkeit und Angemessenheit maßgeblich. Bezüglich der Erforderlichkeit ist bei der Reaktion auf Cyberangriffe insbesondere zu bedenken, ob ein milderes Mittel, also beispielsweise rein defensive Maßnahmen wie Firewalls oder Ähnliches ausreichend sind.⁸⁷ Stehen im konkreten Fall keine milderen Mittel zur Verfügung, sind – unter Wahrung der Angemessenheit – durchaus auch offensive Gegenmaßnahmen denkbar. Diese Maßnahmen können nach dem Wortlaut von Art. 51 UN-Charta wie bei konventionellen Angriffen einzeln oder auch kollektiv, also zum Beispiel auch im Rahmen der NATO, getroffen werden.⁸⁸ Zu beachten ist allerdings, dass die Beweislast sowohl für die Gegenwärtigkeit als auch für die Verhältnismäßigkeit bei dem Staat liegt, der sich auf das Selbstverteidigungsrecht beruft.⁸⁹

Als letzte Grenze des Selbstverteidigungsrechts ist schließlich die Anzeigepflicht gegenüber dem UN-Sicherheitsrat zu beachten. Dies ist eine echte Rechtspflicht und sichert das Monopol, welches dem Sicherheitsrat zur Wahrung des Weltfriedens zukommt. Das Selbstverteidigungsrecht endet nach dem klaren Wortlaut von Art. 51 UN-Charta, sobald der Sicherheitsrat die zur Wahrung des Weltfriedens und der internationalen Sicherheit erforderlichen Maßnahmen getroffen hat. Allerdings führt ein Unterlassen der Anzeige auch nach dem Internationalen Gerichtshof nicht zur Rechtswidrigkeit einer Selbstverteidigungshandlung.⁹⁰ Obwohl in der Anzeigepflicht damit vor al-

85 v. *Arnauld*, Völkerrecht (Fn. 22), S. 503 Rn. 1097; s.a. *A. Randelzhofer*, in: Simma, Charta der Vereinten Nationen, Kommentar, München 1991, Art. 51 Rn. 37.

86 *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 349 Rule 72, Erläuterung 5.

87 *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 349 Rule 72, Erläuterung 3.

88 Dazu auch *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 354 f. Rule 75 sowie *Focarelli*, Self-defence (Fn. 1), S. 317 (336 f.).

89 IGH, *Oil Platforms* (Iran v. USA), Urteil vom 06.11.2003, ICJ-Reports 2003, 161, §§ 57 u. 76; *A. Randelzhofer/G. Nolte*, in: Simma/Khan/Nolte/Paulus (Fn. 14), Art. 51 Rn. 45 f.

90 IGH, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. USA), Merits, Urteil vom 27.06.1986, ICJ-Reports 1986, 14, § 235; dazu auch *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 356 Rule 75, Erläuterung 1.

lem die Subsidiarität des Selbstverteidigungsrechts zum Ausdruck kommt, bleibt es in der Praxis gleichwohl höchst relevant. Schließlich wird der Sicherheitsrat oftmals blockiert sein oder aus anderen Gründen seine Aufgabe nicht oder jedenfalls nicht innerhalb kurzer Zeit wahrnehmen können.⁹¹ Im aktuellen Ukraine-Krieg zeigt sich diese Blockade besonders deutlich, da Russland nicht nur der Aggressor, sondern auch ständiges Mitglied im UN-Sicherheitsrat mit Vetorecht ist. Der Resolutionsentwurf zur Beendigung des Ukraine-Krieges vom 25.02.2022 scheiterte am russischen Veto.⁹² Dass die Generalversammlung am 02.03.2022 und 24.03.2022 mit zwei Resolutionen das Vorgehen Russlands als völkerrechtswidrige Aggression eingeordnet und den unverzüglichen Abzug der russischen Truppen gefordert hat,⁹³ vermag wegen der fehlenden völkerrechtlichen Verbindlichkeit von Resolutionen der Generalversammlung die Blockade nicht zu überwinden. Gleichwohl ist die Verabschiedung mit einer Mehrheit von 141 bzw. 140 Stimmen eine nicht zu unterschätzende politische Botschaft der Weltgemeinschaft, die sich gegen die russische Aggression stellt.⁹⁴

Ungeachtet einer Blockade dürfte im Allgemeinen, insbesondere bei einzelnen Cyberattacken, bei denen aus den geschilderten Gründen der Faktor Zeit in mehrfacher Hinsicht entscheidend ist, die Rolle des Sicherheitsrats in der Zukunft eher gering bleiben. Wegen der Schnelligkeit von Cyberattacken, erlangt auch die sonst schwierige Frage, ob die vom Sicherheitsrat ergriffenen Maßnahmen auch ausreichend waren,⁹⁵ keine allzu große Bedeutung. Unabhängig von der künftigen Rolle Russlands in der UN dürfte die Rolle des Sicherheitsrats jedenfalls bei einzelnen kurzfristigen Cyberattacken schon rein faktisch eher gering bleiben. In rechtlicher Hinsicht bleibt es aber dabei, dass der Sicherheitsrat nach Art. 51 UN-Charta auch bei Cyberattacken zwingend zu unterrichten ist. Ergreift der Sicherheitsrat in der Folge erfolgreich Maßnahmen zur Friedenssicherung, endet das Selbstverteidigungsrecht. Eine andere Frage ist es, ob der Sicherheitsrat in seiner heutigen Organisation und Zusammensetzung den aktuellen Herausforderungen des Friedenssicherungsrechts gerade mit Blick auf Cyberspace noch gerecht werden kann.

91 Siehe nur *Randelzhofer*, in: *Simma* (Fn. 85), Art. 51 Rn. 36.

92 UN-SC, Pressemitteilung SC/14808 vom 25.02.2022, abrufbar unter: <https://www.un.org/press/en/2022/sc14808.doc.htm> (zuletzt abgerufen: 23.02.2023).

93 UN Generalversammlung, A/RES/ES-11/1 vom 02.03.2022 und A/RES/ES-11/2 vom 24.3.2022.

94 So auch *Schmahl*, *Angriffskrieg* (Fn. 4), 971.

95 Siehe nur *Stein/v. Buttlar/Kotzur*, *Völkerrecht* (Fn. 38), S. 297 Rn. 798.

C. Zusammenfassung und Ausblick

An dieser Stelle soll es aber nicht um die Zukunft der Friedenssicherung im Rahmen der UN und die künftige Rolle Russlands im internationalen System gehen. Stattdessen sollen zum Abschluss noch einmal die wesentlichen Erkenntnisse zusammengefasst und ein kleiner Ausblick auf künftige Entwicklungen im Zusammenspiel von Cyberspace und Völkerrecht gewagt werden.

Cyberspace stellt das klassische Völkerrecht wegen seiner weltweiten netzwerkartigen Struktur und seiner Ungebundenheit an einen bestimmten Standort vor nicht unerhebliche Herausforderungen. Gleichwohl – und das ist der erste wichtige Punkt – führt dies nicht dazu, dass Cyberspace ein rechtsfreier Raum wäre. Vielmehr können die Regeln des Völkerrechts auch auf Cyberattacken angewendet werden. Cyberangriffe können und sollten aber nicht pauschal in eine bestimmte völkerrechtliche Kategorie eingeordnet werden. Vielmehr ist entsprechend der Abstufungen der UN-Charta im Einzelfall zu prüfen, ob sich eine Cyberattacke als Verstoß gegen das Interventionsverbot, als unzulässige Gewaltanwendung oder als bewaffneter Angriff einordnen lässt.⁹⁶ Wie dargestellt, können Cyberangriffe einen „bewaffneten Angriff“ im Sinne des Art. 51 UN-Charta darstellen, wenn sie in Ausmaß („*scale*“) und Wirkung („*effect*“) mit einem konventionellen Schlag vergleichbar sind. Unter strenger Achtung von Unmittelbarkeit und Verhältnismäßigkeit ist dann eine Selbstverteidigung mit digitalen oder konventionellen Mitteln möglich. Ist die Schwelle eines „bewaffneten Angriffs“ nicht erfüllt, ist dem betroffenen Staat das Selbstverteidigungsrecht verwehrt. Wegen des regelmäßig durch die feindliche Cyberoperation vorliegenden Verstoßes gegen das Interventionsverbot ist es dem Staat aber unbenommen, friedliche völkerrechtliche Gegenmaßnahmen zu ergreifen.

In beiden Fällen steht der Staat jedoch regelmäßig vor dem Problem der schwierigen bis unmöglichen Rückverfolgbarkeit von Cyberoperationen. Die Antwort auf die Frage, ob und wie völkerrechtliche Instrumente zur Überwindung dieser in erster Linie technischen Herausforderung genutzt werden können, ist eine der großen Aufgaben für die Weiterentwicklung des Völkerrechts. Einige der Ansätze, die in diesem Zusammenhang diskutiert werden, sollen zum Abschluss vorgestellt werden.

96 Hobe, Völkerrecht (Fn. 13), S. 247; Y. Dinstein, Computer Network Attacks and Self-Defense, *International Law Studies* 76 (2002), 99 ff.

Zu diesem Zweck bietet es sich an, noch einmal auf den geschilderten Fall zurückzukommen, bei dem ein Cyberangriff zweifelsfrei von Computern oder Servern eines bestimmten Staates verübt wurde, aber die handelnde Person entweder gar nicht zu identifizieren oder aber ihr Verhältnis zum Staat nicht eindeutig ist. Ist die Person zu identifizieren, dürfte gleichwohl eine Zurechnung zum jeweiligen Staat schwer möglich sein. Die nach der bisherigen Rechtsprechung geforderte „Entsendung“ oder „effective control“ wird sich regelmäßig nicht ohne Weiteres nachweisen lassen, zumal der betroffene Staat den Cyberangriff kaum offen zugeben wird. Eine Möglichkeit wäre nun, in Bezug auf Cyberangriffe die Zurechnungskriterien generell zu lockern. Ähnliches wäre auch bei den Grenzen des Selbstverteidigungsrechts denkbar. Beispielsweise könnte das Gegenwärtigkeitskriterium mit Blick auf den Zeitfaktor bei Cyberangriffen großzügiger gehandhabt werden. Allerdings ist bei jeder extensiven Auslegung des Selbstverteidigungsrechts die mögliche Missbrauchsgefahr mitzubedenken. Die Grundidee der Nachkriegsordnung durch die Vereinten Nationen ist die Sicherung des Weltfriedens durch ein möglichst weitreichendes Gewaltverbot.⁹⁷ Vor diesem Hintergrund sollten Modifikationen der Tatbestandsvoraussetzungen oder Zurechnungskriterien nur mit äußerster Vorsicht vorgenommen werden.⁹⁸

Diskutiert wird in diesem Zusammenhang insbesondere eine Modifikation der Zurechnung im Wege einer Beweislastumkehr. Vor dem Hintergrund, dass die klassischen Zurechnungskriterien im Recht der Staatenverantwortlichkeit für die deliktische Haftung geschaffen wurden, könnte man tatsächlich überlegen, eine derartige Modifikation für jene Fälle anzunehmen, bei denen der Ursprung der Cyberattacke zweifelsfrei einem Staatsgebiet zugeordnet werden kann. Da es bei Cyberoperationen bzw. deren Abwehr weniger um eine *Haftung für* als um eine *Prävention von* Schäden geht, wäre es durchaus denkbar, die Zurechnung in solchen Szenarien widerleglich zu vermuten. Konkret würde dies bedeuten, dass ein Cyberangriff widerleglich dem Staat zugerechnet wird, von dessen Territorium der Angriff ausgeht.⁹⁹

97 Vgl. Art. 1 Nr. 1 und 4, Art. 2 Nr. 4 und Art. 103 UN-Charta; zur historischen Entwicklung siehe nur *Hobe*, Völkerrecht (Fn. 13), S. 26 ff. u. S. 205 ff.

98 So auch *S. Pangrazzi*, Self-defence against Cyberattacks?, Digital and kinetic defence in light of article 51 UN-Charter, Policy Brief, ICT for Peace Foundation, Genf 2021, S. 19 ff., abrufbar unter: <https://ict4peace.org/wp-content/uploads/2021/03/ICT4Peace-2021-Cyberattacks-and-Article51-1.pdf> (zuletzt abgerufen: 23.02.2024); s.a. *Krieger*, anonymous (Fn. 50), 1 (11 ff.).

99 *Heintschel von Heinegg*, Cyberspace (Fn. 21), S. 159 (172); s.a. *Schaller*, Internationale Sicherheit (Fn. 70), S. 23 ff. m.w.N.

Der Staat würde – vorläufig – als verantwortlich für den Angriff angesehen, weil er den Missbrauch seines Netzwerks geduldet bzw. nicht verhindert hat. Ein Ansatzpunkt für ein Selbstverteidigungsrecht gegenüber diesem Staat wäre die grundsätzliche Pflicht jedes Staates, dafür zu sorgen, dass von seinem Staatsgebiet keine Angriffe auf andere Staaten ausgehen.¹⁰⁰ Eine Selbstverteidigung wäre aber regelmäßig auch dann nur *erforderlich*, wenn der Staat, von dessen Territorium der Angriff ausging, nicht willens oder in der Lage war gegen die privaten Akteure vorzugehen. Aus diesem Grund müsste dem betreffenden Staat zunächst die Möglichkeit eröffnet werden, sich zu erklären und selbst geeignete Maßnahmen durchzuführen.¹⁰¹ Aufgrund ihrer Widerlegbarkeit ist die Vermutung im Ergebnis aber eben nicht mehr als eine Umkehr der Beweislast. Darüber hinaus ist eine derartige Umkehr der Beweislast nicht nur im Hinblick auf ihre Vereinbarkeit mit den grundsätzlichen Regeln der Staatenverantwortlichkeit,¹⁰² sondern auch und vor allem im Hinblick auf eine drohende Eskalation bedenklich.¹⁰³ Schließlich hilft eine Beweislastumkehr nicht weiter, wenn der Ursprung eines Angriffs gar nicht einem bestimmten Staatsterritorium zugeordnet werden kann.¹⁰⁴

Das Problem, wie mit dem Umstand der fehlenden Kenntnis über den Ursprung von Cyberattacken umgegangen werden soll, bleibt also weiterhin bestehen. Ein vielversprechender Ansatz könnte hier in der Orientierung am Umweltvölkerrecht liegen, welches ebenfalls mit dem Problem des „Nicht-Wissens“ umgehen muss. Im Umweltrecht wird mit guten Gründen das Vorsorgeprinzip (sog. „*precautionary principle*“) zur Anwendung gebracht. Demnach müssen Aktivitäten, die schädlich für die Umwelt sind, vermieden und Maßnahmen zur Prävention getroffen werden, und zwar gerade auch dann, wenn Ungewissheit über Kausalität und Auswirkungen der möglicher-

100 Vgl. IGH, *Corfu Channel Case* (United Kingdom v. Albania), Procedures, Urteil vom 25.03.1948, ICJ-Reports 1948, 15; s.a. *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 30 ff. Rule 6 und S. 43 ff. Rule 7 sowie *K. Ziolkowski*, General Principles of International Law as Applicable in Cyberspace, in: *K. Ziolkowski* (Hrsg.), *Peacetime Regime for State Activities in Cyberspace: International Law, International Relations and Diplomacy*, Tallinn 2013, S. 135, 165 ff. u. 185 ff.

101 *Kolossa*, Cyberraum (Fn. 66), 167; s.a. *Schmitt* (Hrsg.), Tallinn Manual 2.0 (Fn. 52), S. 347 Rule 71, Erläuterung 25.

102 Insoweit kritisch *C. Droege*, Get off my cloud: cyber warfare, international humanitarian law, and the protection of civilians, IRCC 94 (2012), 533 (543 f.).

103 Ebenso *Dornbusch*, Kampfführungsrecht (Fn. 1), S. 131 f. u. 229.

104 *Heintschel von Heinegg*, Cyberspace (Fn. 21), S. 159 (172).

weise schädlichen Handlungen besteht.¹⁰⁵ Interessant ist in diesem Zusammenhang Punkt 15 der rechtlich nicht bindenden Rio-Deklaration von 1992, der ausdrücklich dazu auffordert, wissenschaftliche Unsicherheit („*scientific uncertainty*“) nicht als Vorwand für das Unterlassen wesentlicher präventiver Maßnahmen zur Verhinderung potenzieller zukünftiger Umweltschäden zu nehmen.¹⁰⁶

Dieser Grundgedanke könnte auch für Cyberoperationen fruchtbar gemacht werden. Nicht mehr der Angriffspunkt wäre entscheidend für die Zuordnung einer Cyberattacke zu einem Aggressor-Staat, sondern es käme im Wesentlichen auf den Schutz vor den mit ihr einhergehenden Auswirkungen an. Vorsorge könnte im internationalen Bereich ein Verbot möglicher schädigender Handlungen zum Inhalt haben, was etwa in ein umfassendes Entwicklungs-, Herstellungs- und Nutzungsverbot für Cyberwaffen münden könnte.¹⁰⁷ Für die Staaten selbst könnte dies allerdings auch bedeuten, dass sie in ihrem eigenen regulativen Netzwerk sicherstellen müssten, dass entsprechende Missbrauchshandlungen des Cyberspace ausgeschlossen würden. Auf völkerrechtlicher Ebene ließe sich insofern an die Etablierung eines Kodex von jedenfalls (halb)verbindlichen Grundsätzen zur Erhöhung der Cybersicherheit denken, deren Einbezug Staaten nur durch strikte Regeln innerstaatlicher Rechtssetzung Folge leisten könnten. Da weder nach national- noch nach internationalem Recht die letztlich technischen Probleme der Rückverfolgung vollständig lösbar sind, erscheint ein derartiger Perspektivwechsel von der Repression auf die Prävention besonders vielversprechend.

105 Ausf. *Schulze*, Cyber-„War“ (Fn. 6), S. 212 ff. sowie *T. Marauhn*, in: K. Ziolkowski (Hrsg.), *Peacetime Regime for State Activities in Cyberspace: International Law, International Relations and Diplomacy*, Tallinn 2013, S. 465 ff., insb. S. 474 ff.

106 Rio Deklaration über Umwelt und Entwicklung vom 14.06.1992, ILM 31 (1992), 876.

107 Dazu und zum Folgenden auch ausf. *Schulze*, Cyber-„War“ (Fn. 6), S. 216 ff.; s.a. *Krieger*, anonymous, 1 (4 ff.).

Der Einfluss von Social Bots auf Debatten in sozialen Netzwerken am Beispiel Twitter

Amina Hoppe

A. Einleitung und Begriffsklärungen

In den Medien wird immer wieder berichtet, dass (Social) Bots maßgeblich die Debatten in sozialen Netzwerken bestimmen würden. Schlagzeilen legen nahe, dass diese automatisierten Accounts Meinungsbilder verzerren, Propaganda betreiben und sogar Wahlen entscheiden könnten. Der vorliegende Beitrag möchte diesen Mythos des menschengleichen Social Bots genauer beleuchten und klären, wie Bots funktionieren und wie leistungsfähig sie tatsächlich sind. Dabei sollen auch wissenschaftliche Studien genauer beleuchtet und ihre Aussagekraft hinterfragt werden.

Unter einem Bot soll dabei zunächst ein Programm oder automatisiertes Skript (Algorithmus) verstanden werden, das einmalig oder dauerhaft die ihm vordefinierten Aufgaben selbstständig abarbeitet. Von einem Social Bot im Konkreten spricht man bei einem maschinell betriebenen Account eines sozialen Netzwerks, dem eine algorithmische Automatisierung zugrunde liegt und der (mehr oder weniger) vorgibt, ein(e) menschliche(r) Nutzer*in zu sein.¹ Die verschiedenen Definitionen weichen dabei aber bereits voneinander ab, was die Einordnung von Forschungsergebnissen erschwert.² Auch enthält der Begriff keine Wertung über die Nützlichkeit oder Schädlichkeit des Algorithmus.³ So können Social Bots auch nützliche Informationen zu einem Thema zusammentragen oder im Sinne der Barrierefreiheit Inhalte anderer Nutzer*innen auswerten.

1 A. Thieltes/S. Hegelich, Falschinformationen und Manipulation durch social bots in sozialen Netzwerken, in: A. Blätte/J. Behnke/K.-U. Schnapp/C. Wagemann (Hrsg.), *Computational Social Science*, Baden-Baden 2018, S. 357 (357, 359); E. Ferrara/O. Varol/C. Davis/F. Menczer/A. Flammini, The rise of social bots, *Commun. ACM* 2016, 96 (96).

2 Zum Begriff auch Ch. Grimme/M. Preuss/L. Adam/H. Trautmann, Social Bots: Human-Like by Means of Human Control?, <https://arxiv.org/pdf/1706.07624.pdf> (zuletzt abgerufen am 30.01.2023).

3 Thieltes/Hegelich, Falschinformationen (Fn. 1), S. 358 f.

Soziale Netzwerke sind dabei solche Onlinedienste, die die Möglichkeit zu Informationsaustausch und Beziehungsaufbau bieten. Da die meisten Studien zur Wirkung von Bots auf das Netzwerk Twitter fokussiert sind und auch die mediale Aufmerksamkeit sich damit auf dieses Netzwerk konzentriert, will es auch dieser Beitrag tun.

Bei Twitter handelt es sich um ein soziales Netzwerk, das 2006 gegründet wurde und vorrangig Beiträge mit bis zu 280 Zeichen Länge erlaubt. Diese können im Regelfall von allen Internetnutzer*innen, auch ohne Anmeldung, gelesen werden. Followern des verfassenden Accounts werden die Beiträge auch in einer sog. „Timeline“ angezeigt, ebenso denjenigen Accounts, die anderen Accounts folgen, die auf einen Beitrag antworten, ihn teilen (retweeten) oder mögen („Gefällt mir“). In Deutschland nutzen etwa 2 % der Menschen das Netzwerk und tauschen sich dabei zu vielfältigen Themen aus, insbesondere zu Politik und aktuellen Nachrichten. Verglichen mit anderen Netzwerken dürften Twitter und Facebook als die politischsten sozialen Netzwerke in Deutschland gelten.⁴

B. Die Funktionsweise von social bots

Um zu erkennen, welche Möglichkeiten der Einsatz von (Social) Bots überhaupt bietet, sollte der Blick zunächst darauf gerichtet werden, welche Fähigkeiten Bots haben. Dies reicht von ganz einfachen Programmierungen bis hin zu sehr komplexen Funktionen. Die Programmierung richtet sich dabei vor allem nach der jeweiligen Zielsetzung.⁵ Gleichzeitig finden sich verschiedene Generationen von Bots mit unterschiedlicher Qualität.⁶ Daher und aufgrund der Schwierigkeiten der Detektion ist das wirkliche Potenzial dieser Algorithmen nur schwer einzuschätzen.⁷

Basisfunktionen von Social Bots auf Twitter können dabei das Posten von zuvor manuell erstellten Beiträgen (Tweets) bzw. von bestimmten Links,⁸

4 Ähnlich wirkt der Messengerdienst Telegram, bei dem es sich jedoch nicht um ein soziales Netzwerk handelt.

5 Thielges/Hegelich, Falschinformationen (Fn. 1), S. 367.

6 Thielges/Hegelich, Falschinformationen (Fn. 1), S. 367.

7 L. M. Aiello/M. Deplano/R. Schifanella/G. Ruffo, People are Strange when you're a Stranger: Impact and Influence of Bots on Social Networks, in: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, Dublin 2012, S. 10.

8 Nach einer Studie des Pew Research Centers werden Links sogar überproportional häufig von Bots gepostet, vgl. St. Wojcik/S. Messing/A. Smith/L. Rainie/P. Hitlin, Bots in the

ein einfaches Teilen/Retweeten fremder Beiträge (z. B. aller Tweets mit einem bestimmten Hashtag), die Auswertung des Netzwerks oder auch Freundschaftsanfragen sein. Etwas komplexer hingegen ist das intelligente Retweeten (z. B. aller Posts, die eine bestimmte Meinung vertreten), das Erstellen eigener Inhalte mittels KI, das inhaltlich korrekte Kommentieren und Antworten auf Beiträge von Dritten, also eine thematisch passende Beteiligung an Diskussionen/Konversationen, oder auch Recherchen im Internet vor dem Erstellen von Beiträgen.⁹ Eine Studie von *Assenmacher et. al.* zeigt jedoch deutlich, dass derart intelligente Bots kaum im Internet, weder in öffentlichen Repositorien noch im Darknet, verfügbar sind, also nicht ohne Weiteres von Nutzer*innen mit geringen Kenntnissen genutzt werden können.¹⁰ Die Autor*innen stellen auch klar, dass es vermutlich einfacher und Kosten-Nutzen-effizienter ist, einzelne Personen mit der Betreuung von Fake-Accounts zu beauftragen.¹¹ Es ist anzunehmen, dass derartige hybride Formen mit unterschiedlichen Graden an Automatisierung ebenfalls häufig in sozialen Netzwerken anzutreffen sind.

Besonders interessant sind diejenigen Eigenschaften der Bots, die das menschliche Verhalten nachahmen sollen (Mimikry).¹² So kopieren Bots den menschlichen Tagesablauf, bauen Tippfehler in die von ihnen geposteten Beiträge ein oder verfassen themenfremde Nachrichten, um ihren eigentlichen Einsatzzweck zu verschleiern.¹³ Eben diese Nachahmung macht es auch besonders schwer, von Bots verfasste Beiträge sicher zu erkennen, da hier gerade versucht wird, den automatisierten Ursprung zu verschleiern.

Twittersphere, <https://www.pewresearch.org/internet/2018/04/09/bots-in-the-twitter-sphere> (zuletzt abgerufen am 30.11.2022).

- 9 *Ferrara/Varol/Davis/Menczer/Flammini*, The rise of social bots (Fn. 1), 99.
- 10 *D. Assenmacher/L. Clever/L. Frischlich/Th. Quandt/H. Trautmann/Ch. Grimme*, Demystifying Social Bots: On the Intelligence of Automated Social Media Actors, in: *Social Media and Society*, July-September 2020, 1 (10).
- 11 *Assenmacher/Clever/Frischlich/Quandt/Trautmann/Grimme*, Demystifying Social Bots (Fn. 10), II. Auch *Grimme/Preuss/Adam/Trautmann*, Social Bots (Fn. 2), 5 zeigen, dass es relativ einfach ist, derartige Verhaltensmuster für Bots zu programmieren, die Generierung von geeigneten Inhalten aber ungleich schwerer ist.
- 12 *S. Hegelich/D. Janetzko*, Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet, in: *Tenth International AAAI Conference on Web and Social Media (ICWSM) 2016*, 579 (582).
- 13 *S. A. Golder/M. W. Macy*, Diurnal and seasonal mood vary with work, sleep, and day-length across diverse cultures, *Science* 2011, 1878; *Ch. Grimme/D. Assenmacher/L. Adam*, Changing Perspectives: Is It Sufficient to Detect Social Bots?, *Social Computing and Social Media. User Experience and Behavior*, Juli 2018, S. 445 (447) m. w. N.

C. Einsatzgebiete von social bots

Mit diesen Fähigkeiten können Bots in sozialen Netzwerken vielfältig eingesetzt werden. Im Folgenden soll der Fokus auf schädigendem Verhalten liegen. Die zahlreichen positiven Nutzen dieser Algorithmen, wie Zusammenfassungen von Tweets, Verbesserung der Barrierefreiheit und ähnliches, sollen jedoch nicht unerwähnt bleiben.

Besonders relevant bei der Betrachtung der verschiedenen Einsatzbereiche sind insbesondere sog. Botnets (Botnetze), bei denen mehrere automatisiert betriebene Accounts miteinander vernetzt sind. Diese werden auf einem oder mehreren Rechnern ausgeführt und von einer Person gelenkt.¹⁴ Hierdurch kann der Eindruck vermittelt werden, es handele sich um eine Vielzahl an (realen) Nutzern, die die gleiche Meinung zu einem Thema vertreten.¹⁵ Hier besteht eine gewisse Ähnlichkeit zu sog. Trollarmeen. „Trolle“ können Bots sein, im Regelfall werden die entsprechenden Accounts aber von Menschen nach bestimmten Vorgaben geführt.¹⁶

Am offensichtlichsten ist wohl die Fähigkeit von Bots, in die Kommunikation, bei massenweisem Auftreten auch in die Kommunikationskultur bei bestimmten Debatten oder sogar des gesamten sozialen Netzwerks einzugreifen. So können gezielt Interaktionen (etwa durch Spam) gestört werden.¹⁷ Das Posten beleidigender oder ähnlich negativer Inhalte kann darüber hinaus zu einer Vergiftung des Kommunikationsklimas beitragen.

Daneben können Bots jedoch auch inhaltlich in Debatten eingreifen, was insbesondere bei gesellschaftlichen oder politischen Themen von Belang ist. So können sie etwa für den Wahlkampf oder als außenpolitische Strategie genutzt werden.¹⁸ Darüber hinaus können Bots dazu dienen, massenhaft (Falsch-)Informationen zu verbreiten, die insbesondere eine bestimmte politische Auffassung stützen oder Hass verbreiten sollen. Dies kann sich bei den sog. social hoaxes auf mehrere soziale Netzwerke erstrecken, um die Glaubwürdigkeit noch weiter zu erhöhen, Verwirrung zu stiften und ggf.

14 Thieltges/Hegelich, Falschinformationen (Fn. 1), S. 359.

15 P. Egli/D. Rechsteiner, Social Bots und Meinungsbildung in der Demokratie, *AJPPJA* 2017, 249 (250).

16 A. Dawson/M. Innes, The Internet Research Agency in Europe 2014–2016, https://www.cardiff.ac.uk/_data/assets/pdf_file/0004/1490548/CSRI-IRA-Report-Final.pdf (zuletzt abgerufen am 30.11.2022).

17 Thieltges/Hegelich, Falschinformationen (Fn. 1), S. 357.

18 Thieltges/Hegelich, Falschinformationen (Fn. 1), S. 362 m. w. N.

Panik und Hysterie auszulösen.¹⁹ Aber auch das Zustimmung oder Verbreiten von bestimmten, von anderen Accounts verbreiteten Meinungen durch Bots kann in Debatten eingreifen.²⁰ Wie Studien gezeigt haben, kann allein eine gewisse Anzahl an Fake Followern einen Einfluss auf die Bewertung eines Accounts haben, der durch das hohe Maß angeblicher Vernetzung beliebter und einflussreicher erscheint, was durch die Algorithmen der Netzwerke noch verstärkt werden kann.²¹ Diese Bots können einzelne Beiträge teilen und/oder retweeten. Durch ihre Automatisierung und die Möglichkeit, viele Accounts gleichzeitig zu lenken,²² können Bots zudem sog. Graswurzelbewegungen²³ vorspielen oder die Ergebnisse einer Suchmaschinen-Suche beeinflussen (sog. „Twitter Bombs“).²⁴ Darüber hinaus können Bots (Falsch-)Meldungen im sozialen Netzwerk aufrechterhalten, indem sie, einmal programmiert, immer wieder Inhalte hierzu posten und alte Tweets retweeten.

Unabhängig von gesellschaftlichen wie politischen Themen können Bots weitere Schädigungsabsichten verfolgen, etwa Werbung, gefälschte Produkte oder Produktbewertungen in das soziale Netzwerk einschleusen. Auch Malware oder Phishing-Software kann durch Bots gezielt verbreitet werden.²⁵

D. Der Einfluss auf Debatten

Nach dieser Einführung soll nun auf den (vermeintlichen) Einfluss von Bots auf Debatten in sozialen Netzwerken eingegangen werden.

19 *Thieltges/Hegelich*, Falschinformationen (Fn. 1), S. 363.

20 *Thieltges/Hegelich*, Falschinformationen (Fn. 1), S. 360; *A. Bessi/E. Ferrara*, Social bots distort the 2016 U.S. Presidential election online discussion, <https://firstmonday.org/ojs/index.php/fm/article/view/7090/5653> (zuletzt abgerufen am 30.11.2022).

21 *Thieltges/Hegelich*, Falschinformationen (Fn. 1), S. 361.

22 *Thieltges/Hegelich*, Falschinformationen (Fn. 1), S. 358.

23 Vgl. hierzu etwa die Studie von *J. Ratkiewicz/M. Conover/M. Meiss/B. Concalves/A. Flammini/F. Menczer*, Detecting and Tracking Political Abuse in Social Media, <https://ojs.aaai.org/index.php/ICWSM/article/view/14127> (zuletzt abgerufen am 30.11.2022), S. 297.

24 *P. T. Metaxas/E. Mustafaraj*, From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search, <https://repository.wellesley.edu/object/ir122> (zuletzt abgerufen am 30.11.2022), S. 5.

25 *Thieltges/Hegelich*, Falschinformationen (Fn. 1), S. 363.

Es ist davon auszugehen, dass etwa 15 % der Accounts auf Twitter Social Bots sind.²⁶ Wie gerade beschrieben, tragen diese nicht nur in Einzelbeiträgen Inhalte zu Debatten bei, sondern können im „Zusammenwirken“, also durch die Nutzung mehrerer Bots, den Eindruck einer „Meinungsübermacht“ erwecken und ihren bereits für sich genommen schädlichen Einfluss noch verstärken. Durch das Liken und (Re-)Tweeten bestimmter Inhalte können sie darüber hinaus auch Trends setzen, die häufig vom Netzwerk aufgegriffen werden. Wie sehr dies die Debatten bestimmt, ist wissenschaftlich allerdings umstritten. Wie die eingangs erwähnten Medienberichte suggerieren und einige Studien nahelegen, sollen Bots jedoch einen großen Einfluss haben können.

So titelte etwa Bloomberg „Twitter Bots Helped Trump and Brexit Win, Economic Study Says“.²⁷ Dieser Schlagzeile liegt die Studie von *Gorodnichenko et al.*²⁸ zu Grunde.²⁹ Diese untersuchte die Informationsflüsse und Reaktionen verschiedener Twitter-Accounts und zog daraus Rückschlüsse auf das Wählerverhalten. Sie kam zu dem Ergebnis, dass Social Bots das Ergebnis des Brexit-Referendums um 1,76% zugunsten eines Austritts und die US-Wahl 2016 um 3,23% zugunsten Trumps verschoben hätten. Dem lag die Annahme zu Grunde, dass die Aktivität auf Twitter mit dem Wahlausgang korreliere, was aufgrund der geringen Nutzerzahlen so wohl schon im Ansatz nicht auf Deutschland übertragbar wäre. Unabhängig von der undurchsichtigen Berechnung der angegebenen Prozentwerte überraschen allerdings vor allem die Kriterien, nach denen festgestellt wurde, ob es sich bei einem Account um einen Bot handelt. Ein Account wurde nämlich dann als Bot eingestuft, wenn er an mehr als 50 % der untersuchten Tage eines der folgenden Kriterien erfüllte: (a) 10 Tweets (Brexit) bzw. 15 Tweets (US-Wahl) pro Tag, (b) 5 Tweets (Brexit) bzw. 10 Tweets (US-Wahl) zwischen 0 und 6 Uhr, (c) 3 Tweets mit gleichem Inhalt an einem Tag oder (d) „neu erstellter Account“ (nach Bekanntgabe des Referendums bzw. des Vizepräsidentchaftskandida-

26 Ferrara/ Varol/Davis/Menczer/Flammini, The rise of social bots (Fn. 1), 96.

27 J. Smialek, Twitter Bots Helped Trump and Brexit Win, Economic Study Says, <https://www.bloomberg.com/news/articles/2018-05-21/twitter-bots-helped-trump-and-brexit-win-economic-study-says> (zuletzt abgerufen am 30.11.2022).

28 Y. Gorodnichenko/T. Pham/O. Talavera (2018): Social media, sentiment and public opinions: Evidence from #Brexit and #USElection, https://www.nber.org/system/files/working_papers/w24631/w24631.pdf (zuletzt abgerufen am 30.11.2022).

29 Vgl. zur Kritik auch F. Gallwitz/M. Kreil, The Rise and Fall of ‘Social Bot’ Research, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3814191 (zuletzt abgerufen am 30.11.2022), S. 5.

ten). Was aber sagen diese Kriterien über den Automatisierungsgrad des Accounts aus? Letztendlich bestimmt der Algorithmus der Studie allein das Level der Aktivität eines Accounts. Die genannten Mengen an Tweets könnten ohne Weiteres auch von besonders aktiven Nutzer*innen, etwa politischen Aktivist*innen und Wahlhelfer*innen³⁰, verfasst worden sein. Während die Anzahl der Beiträge pro Tag jedoch nach Forschungen zumindest ein Indiz für einen automatisierten Account sein kann (der freilich für sich genommen noch keine validen Studienergebnisse begründen kann), ist insbesondere das letzte Kriterium völlig verfehlt: Es spricht allen Nutzer*innen, die sich neu im Netzwerk engagieren und sich dabei politisch äußern, ihre (menschliche) Existenz ab. Somit sind sämtliche Kriterien, die die Studie an Bots stellt, schlichtweg ungeeignet, diese tatsächlich zu identifizieren. Die Rate sog. falsch-positiver Ergebnisse, also Accounts, die zu Unrecht als Bots gewertet werden, ist aus diesem Grund relativ hoch und die zuvor zitierte, so vermeintlich klar formulierte Schlagzeile bei genauerer Betrachtung mit ganz erheblichen Unsicherheiten zu lesen.

Auch eine Studie von *Howard und Kollanyi*³¹ mit ähnlichen Ergebnissen wurde von den Medien aufgegriffen. Hier besteht allerdings das gleiche Problem wie oben, nur, dass die Studie noch weiter vereinfacht und alle Accounts, die mehr als 50 Tweets (inkl. Retweets) pro Tag posten, als Bots klassifiziert. Dieses sog. „Oxford-Kriterium“, das häufig in Studien genutzt wird, ist jedoch ebenso unpräzise und hat eine hohe Rate falsch-positiver Ergebnisse,³² worauf aber weder die Studie noch die Medien überhaupt nur hinweisen. Auch die mit einem massiven Medienecho verbundene Studie des privaten Unternehmens Botswatch zum Migrationspakt soll diese Regel angewendet haben,³³ beruft sich aber – von den Medien zunächst völlig unbeachtet – hinsichtlich der nicht veröffentlichten Detektionsalgorithmen auf ein Geschäftsgeheimnis.³⁴ Dies muss wohl als Höhepunkt der unkritischen Berichterstattung über derartige Studien angesehen werden.

30 *Gallwitz/Kreil*, Rise and Fall (Fn. 29), S. 3.

31 *P. N. Howard/B. Kollanyi*, Bots, #strongerin, and #brexit: Computational propaganda during the uk-eu referendum, <https://www.ssrn.com/abstract=2798311> (zuletzt abgerufen am 30.11.2022).

32 *Gallwitz/Kreil*, Rise and Fall (Fn. 29), S. 3.

33 *MB*, Mächten Twitter-Bots Stimmung gegen Migrationspakt?, <https://weischer.media/de/de/trends-und-innovationen/blickwinkel/machten-twitter-bots-stimmung-gegen-migrationspakt> (zuletzt abgerufen am 30.11.2022).

34 *R. Tusch*, Kritik an Botswatch: Warum die Debatte um die Social Bot-Studie zum Migrationspakt für Medien wichtig ist, <https://meedia.de/2018/12/12/kritik-an-botswat>

Ein auf den ersten Blick methodisch besseres Vorgehen hat die Studie von *Keller und Klinger*³⁵. Sie untersuchten die Follower der Accounts von sieben deutschen Parteien vor und im Wahlkampf (insgesamt 838.026 Accounts). Dabei wuchs der Anteil der Beiträge von Social Bots an der Debatte während des Wahlkampfes von 7,1 % auf 9,9 % (entspricht etwa 83.000 Bots). Die Klassifizierung als Bot erfolgte dabei nicht mittels einfacher Kriterien, sondern bediente sich eines Algorithmus, der die Wahrscheinlichkeit ausgibt, dass es sich bei dem Account um einen Bot handelt, das sog. Botometer³⁶ der Forschergruppe um *Ferrera*. Dabei handelt es sich um das am häufigsten genutzte Tool zur schnellen Analyse von Bots auf Twitter. Das Botometer folgt der Grundannahme, dass alle Bots bestimmte Eigenschaften haben und Verhaltensweisen an den Tag legen, die sie von echten Nutzer*innen unterscheiden. So sollen Bots mehr retweeten und weniger originäre Inhalte verbreiten, die Accounts sind jünger, antworten seltener und haben längere Usernamen.³⁷ Die genaue Gewichtung dieser und vieler weiterer Eigenschaften lernt der Algorithmus selbstständig im Wege des maschinengestützten Lernens anhand von Trainingsdatensätzen.

Wie wenig reliabel diese Ergebnisse sind, zeigt jedoch die Überprüfung der Studie durch *Gallwitz und Kreil*.³⁸ Nur 20 Monate nach der Wahl erhoben sie noch einmal die Daten der Studie von *Keller/Klinger*. In dieser vergleichsweise kurzen Zeit war der Anteil vermeintlicher Social Bots auf 51,8 % (270.572 Accounts) angewachsen. Der Grund hierfür waren viele Accounts, die länger inaktiv waren, lediglich einen Beitrag gepostet hatten oder durch häufiges Cross-Posting (von Inhalten anderer Plattformen über den sog. „Tweet“-Button) auffielen. Ähnliche Ergebnisse ergaben sich auch in einer anderen Studie der beiden Autoren, die die Mitglieder des US-Kongresses im April 2019 erforschte und dabei einen Anteil von 47 % Bots aufzeigte. Nachdem die Accounts der Mitglieder des US-Kongresses als Trainingsdaten für das Botometer genutzt worden waren, betrug (im Mai 2019) die Botrate nur

ch-warum-die-debatte-um-die-social-bot-studie-zum-migrationspakt-fuer-medien-wichtig-ist (zuletzt abgerufen am 30.11.2022).

35 T. R. Keller/U. Klinger, Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications, *Political Communication* 36 (2019), 171.

36 <https://botometer.osome.iu.edu/> (zuletzt abgerufen am 30.11.2022).

37 *Ferrara/Varol/Davis/Menczer/Flammini*, The rise of social bots (Fn. 1), 102.

38 *Gallwitz/Kreil*, Rise and Fall (Fn. 29).

noch 0,4 %.³⁹ Auch bei Nutzung des Botometers ist also von einer hohen Anzahl falsch-positiver Ergebnisse auszugehen.⁴⁰

Dieser kurze Überblick zeigt, welche methodischen Schwächen derartige Studien haben. *Gallwitz und Kreil* schreiben hierzu: "The idea that armies of 'social bots' are roaming Twitter was created by a small number of researchers with high ambitions but low scientific standards. It meets all the characteristics of a conspiracy theory."⁴¹ Fakt ist jedoch, dass es Bots unbestritten gibt und diese auch in sozialen Netzwerken und in wichtigen und großen Debatten auftreten. Wie viele Bots dort aber aktiv sind und welchen Einfluss sie haben, kann nicht mit Sicherheit gesagt werden. Hierfür sind beobachtende Studien in den sozialen Netzwerken aufgrund der Unmöglichkeit, Bots eindeutig von realen Nutzer*innen unterscheiden zu können, nicht geeignet.

Daher gibt es inzwischen andere Ansätze für Studiendesigns. Eine ebenfalls von größeren Medien⁴² rezipierte Studie haben *Ross et. al.* durchgeführt.⁴³ Sie simulierten ein soziales Netzwerk mit allen Informationsflüssen und untersuchten dabei die sog. „Schweigespire“, nach der Akteure in einer Debatte schweigen, wenn sie das Gefühl haben, in der Minderheit zu sein. Die Studie kam zu dem Ergebnis, dass ein Anteil von 2 bis 4 % Bots ausreichen würde, um in 2 von 3 Fällen das Meinungsklima zu kippen. Wie viele Bots dabei genau gebraucht werden, sei abhängig von der Position im Netzwerk. Die Studie verkennt jedoch, dass in einer Simulation eines Netzwerks genau genommen alle Akteure Bots sind, denn jeder simulierte Akteur ist im Grunde ein automatisierter, algorithmenbasierter Akteur. Weiterhin nimmt die Studie explizit an, dass Bots wie menschliche Akteure handeln würden. Andere Studien zeigen jedoch, dass gerade Personen, insbesondere mit Schulung,

39 *Gallwitz/Kreil*, Rise and Fall (Fn. 29), S. 7; Ebenfalls durch verschiedene Experimente konnten *Ch. Grimme/D. Assenmacher/L. Adam*, Changing Perspectives (Fn. 13), die Fehleranfälligkeit von Botometer nachweisen.

40 *Gallwitz/Kreil*, Rise and Fall (Fn. 29), S. 7 m. w. N. Vgl. auch die Studie von *A. Rauffleisch/J. Kaiser*, The False positive problem of automatic bot detection in social science research, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241045> (zuletzt abgerufen am 30.11.2022).

41 *Gallwitz/Kreil*, Rise and Fall (Fn. 29), S. 13.

42 Vgl. etwa https://rp-online.de/nrw/hochschulen/schon-wenige-social-bots-reichen-aus-um-stimmung-im-netz-zu-machen_aid-36745241 (13.04.2022) und <https://www.tageschau.de/faktenfinder/hintergrund/social-bots-109.html> (jeweils zuletzt abgerufen am 30.11.2022).

43 *B. Ross/L. Pilz/B. Cabrera/F. Brachten/G. Neubaum/S. Stieglitz*: Are social bots a real threat? *European Journal of Information Systems*, 28 (2019), 394.

sehr gut, teilweise sogar besser als Algorithmen, Bots von menschlichen Nutzer*innen unterscheiden können.⁴⁴ Dies wurde jedoch in der Studie nicht berücksichtigt. Insofern geben die genannten Zahlen wohl eher ein Bild darüber, wie manipulativ handelnde Akteure die Meinung im Netzwerk mitbestimmen können, ohne jedoch etwas darüber auszusagen, ob diese Akteure von Menschen oder durch einen Algorithmus gesteuert werden.

Daneben gibt es noch weitere Ansätze, wie das Einrichten einer „Honeypot trap“⁴⁵, um Bots anzulocken, oder die gezielte Interaktion eines für die Studie programmierten Bots mit anderen Nutzer*innen⁴⁶. In derartigen Studien zeigt sich, dass Bots zu zentralen Accounts eines Netzwerks aufsteigen können, wenn sie im Netzwerk aktiv werden, etwa Seiten besuchen oder Kontakt aufnehmen.⁴⁷

Insgesamt erscheint jedoch die Forschungsfrage, wo und wie Bots an Debatten partizipieren, insbesondere aufgrund genereller Erkenntnisschranken wenig ergiebig. Vielmehr sollte eher das „Verhalten“ störender, manipulativer oder sonst negativ agierender Accounts untersucht und (auch) danach gefragt werden, welche Reaktionen auf derartiges Verhalten folgen. Es sollte also letztendlich weniger darum gehen, wie groß der Einfluss von Bots auf Debatten in sozialen Netzwerken ist, sondern vielmehr darum, wie der Einfluss und die gesellschaftlichen Auswirkungen möglichst gering gehalten werden können.

Hierzu können verschiedenste Dinge beitragen. So können Netzwerke durch die Betreiber und durch Gesetze wie den Digital Service Act reguliert werden. Aber auch die Nutzer*innen eines Netzwerks können und sollten

44 G. Wang/M. Mohanlal/Ch. Wilson/X. Wang/M. Metzger/H. Zheng/B. Y. Zhao, Social Turing Tests: Crowdsourcing Sybil Detection, https://gangw.cs.illinois.edu/sybils_ndssl3.pdf (zuletzt abgerufen am 30.11.2022); auch bei den Experimenten von Grimme/Assenmacher/Adam, Changing Perspectives (Fn. 13), S. 456 sind anderen Nutzern die Bots eher aufgefallen als dem Botometer.

45 K. Lee/J. Caverlee/S. Webb, The social honeypot project: Protecting online communities from spammers, <http://portal.acm.org/citation.cfm?doid=1772690.1772843> (zuletzt abgerufen am 30.11.2022).

46 Aiello/Deplano/Schifanella/Ruffo, People are strange (Fn. 7).

47 Aiello/Deplano/Schifanella/Ruffo, People are strange (Fn. 7) zeigten dies in einem kleinen Netzwerk von Bücherliebhabern (aNobii), in dem sie einen Bot installierten, der dort innerhalb von 10 Monaten zu einem der Top-Nutzer aufgestiegen ist und hohe Interaktionsgrade aufwies. Dieser besuchte zunächst lediglich die Benutzerprofile der anderen Accounts, im zweiten Schritt sprach er auch Empfehlungen aus, denen tatsächlich viele Teilnehmer folgten. Grimme/Preuss/Adam/Trautmann, Social Bots (Fn. 2), zeigten die Möglichkeiten der Verbreitung von Bots auf Twitter anhand selbst-programmierter Bots, die jedoch manuell erstellte Inhalte verbreiteten.

eine aktive Rolle einnehmen und Inhalte kritisch prüfen bzw. zur Überprüfung melden. Unabhängig davon, aus welcher Quelle schädigende Beiträge stammen, sollten Nutzer*innen also ein Bewusstsein für Manipulationsversuche entwickeln und für Fake News sensibilisiert werden. Manipulationen, Falschinformationen, Spam und Ähnliches können nicht nur von Bots ausgehen. Und es ist zwar leichter, durch den Einsatz von Bots bestimmte Effekte zu erzielen, diese sind in der Regel allerdings auch leichter erkennbar. Das massenweise Auftreten kann jedoch ebenso durch orchestriertes Vorgehen mehrerer oder durch sog. Trollarmeen hervorgerufen werden, was ebenfalls eine scheinbare „Meinungsübermacht“ bewirkt und zudem in vielen Fällen noch authentischer erscheint. Darüber hinaus sollten Nutzer*innen, aber auch die Medien, sich der Besonderheiten der jeweiligen Netzwerke bewusst sein und hinsichtlich der technischen (Un-)Möglichkeiten sensibilisiert werden, um der Mär von den schädlichen Bots Einhalt zu gebieten.

E. Fazit

Dieser kurze Abriss hat gezeigt, dass Social Bots tatsächlich einen Einfluss auf Debatten in sozialen Netzwerken haben können. Es ist jedoch wissenschaftlich nicht konkret nachweisbar, wie groß dieser Einfluss ist, denn die Detektion von Bots ist nicht ohne Weiteres möglich. Daher sind alle, auch die von den Medien weiterverbreiteten Forschungsergebnisse nicht viel mehr als indizienbasierte Spekulationen. Zudem verlieren eine dahingehende Forschung und eine darauf aufbauende Forderung von Reglementierungen den Blick auf positiv agierende, nützliche Bots, etwa zur Herstellung von Barrierefreiheit. Weit zielführender wäre es hingegen, sich darüber auszutauschen, welchen Einfluss Medien, Netzwerke und deren Nutzer*innen den Bots und anderen Accounts mit schädigendem Verhalten zugestehen wollen. Das Augenmerk sollte entsprechend nicht länger auf der Art der Verbreitung von Inhalten liegen, sondern auf Qualität und Quantität der konkret verbreiteten Inhalte⁴⁸ und ggf. dem orchestrierten Vorgehen.⁴⁹ Auch eine mögliche Regulierung sollte dies beachten. Sie wird am besten gelingen, wenn Nutzer*innen und Netzwerke zusammenarbeiten. Denn um im Sprachspiel der Ringvorlesung zu bleiben: Auch wenn Social Bots in den

48 Grimme/Assenmacher/Adam, *Changing Perspectives* (Fn. 13), S. 456.

49 Für einen Fokus auf orchestriertes Verhalten auch Grimme/Assenmacher/Adam, *Changing Perspectives* (Fn. 13), S. 446.

Amina Hoppe

sozialen Netzwerken das Potenzial zu viel Macht haben, sind Nutzer*innen und Gesellschaft dem nicht machtlos ausgeliefert.

Haftung für Schäden durch KI in der Medizin*

Christian Katzenmeier

A. Künstliche Intelligenz in der Medizin

Künstliche Intelligenz (KI) gehört zu den großen Hoffnungsträgern in der Medizin. Ihr Einsatz verspricht eine bessere individuelle Gesundheitsversorgung durch neue Möglichkeiten der Diagnose und Therapie, der Prävention und der Prädiktion von Krankheiten, sie kann zu einem längeren selbstbestimmten Leben beitragen durch Assistenzsysteme bei der Erledigung von Alltagsaufgaben und Robotik, überdies zu einer Entlastung der Ärzteschaft und der Pflege.¹

KI bezeichnet – bei allen Schwierigkeiten einer begrifflichen Definition – selbständig lernende Software, die in der Lage ist, komplexe Auswahlprozesse unter Einbeziehung einer Vielzahl von Daten autonom durchzuführen und mit Hilfe von softwaregesteuerten Maschinen umzusetzen.² In der Medizin ermöglichen Applikationen auf der Basis intelligenter Datenbanken den Abgleich der individuellen Daten eines Patienten mit einer bestimmten Erkrankung und den Verläufen vergleichbarer Krankheitsfälle einschließlich Nebendiagnosen über Jahrzehnte. Stellten Algorithmen ursprünglich elektronische Auswahlprozesse dar, die von menschlichen Programmierern determiniert waren, so können heute neuronale Netzwerkstrukturen mit Rückkopplungsmechanismen aus großen Datenmengen, mit denen sie gespeist werden, autonom Muster identifizieren, Informationen kategorisieren,

* Aktualisierte Fassung des in MedR 2021, 859–867 publizierten Aufsatzes.

1 C. Katzenmeier, Big Data, E-Health, M-Health, KI und Robotik in der Medizin, MedR 2019, 259.

2 W. Ertel, Grundkurs Künstliche Intelligenz, 3. Aufl., Wiesbaden 2013, S.1 ff.; Bitkom/DFKI (Hrsg.), Künstliche Intelligenz: Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung, Berlin 2017, S. 28; S. Kirn/C. Müller-Hengstenberg, Intelligente (Software-)Agenten: Von der Automatisierung zur Autonomie? Verselbstständigung technischer Systeme, MMR 2014, 225; O. Stiernerling, „Künstliche Intelligenz“ – Automatisierung geistiger Arbeit, Big Data und das Internet der Dinge, CR 2015, 762.

auf diese Weise lernen und auf der Basis wiederum auswählen, d. h. selbst über konkrete Handlungsoptionen entscheiden.³

I. Einsatzgebiete

Aktuell spielt KI eine bedeutsame Rolle, insbesondere in der bildgebenden Diagnostik. Radiomics bezeichnet ein Teilgebiet der medizinischen Bildverarbeitung und radiologischen Grundlagenforschung, die sich mit der Analyse von quantitativen Bildmerkmalen in großen medizinischen Bilddatenbanken befasst.⁴ Hierfür kommt neben Statistik vor allem maschinelles Lernen zum Einsatz. In der Neurologie können KI-gestützte Auswertungen von Gehirnschans einen wichtigen Beitrag zur Früherkennung von Alzheimererkrankungen leisten. In der Kardiologie sind Algorithmen im Einsatz, die Langzeit-EKGs auswerten und Rhythmusstörungen binnen Sekunden finden und auflisten. In der Onkologie können mit Hilfe von KI verschiedene Tumortypen mit großer Treffsicherheit unterschieden werden. KI ist ein zentraler Baustein der Systemmedizin, die das Ziel einer maßgeschneiderten und damit bestwirksamen Therapie verfolgt.⁵ Auch in der Chirurgie kommt KI zum Einsatz. Der derzeit fortschrittlichste chirurgische Roboter *Da Vinci*⁶ ermöglicht komplexe Eingriffe, insbesondere laparoskopische Prostataktomie und Zystektomie, kontrollierter durchzuführen, indem er unwillkürliche Bewegungen des Operateurs ausgleicht. In Zukunft sollen die Roboter darüber hinaus mittels entsprechender KI-Anwendungen Daten aus Voruntersuchungen und aus vorangegangenen Operationen auswerten und

3 H. Dettling/S. Krüger, Digitalisierung, Algorithmisierung und Künstliche Intelligenz im Pharmarecht, PharmR 2018, 513 (514) mit Verweis u.a. auf Bitkom/DFKI (Hrsg.), Künstliche Intelligenz (Fn. 2), S. 32; U. Eberl, Smarte Maschinen, München 2016, S. 100; M. Tegmark, Leben 3.0: Mensch sein im Zeitalter Künstlicher Intelligenz, Berlin 2017, S. 109 ff.; H. Volland, Die kreative Macht der Maschinen, Weinheim 2018, S. 16 f.

4 Informationen unter <https://www.drg.de/de-DE/3601/radiomics/>; zu Herausforderungen und Chancen J. Murray/G. Kaissis/R. Braren/J. Kleesiek, Wie funktioniert Radiomics?, Der Radiologe 2020, 32.

5 Zum Begriff s. D. Hart, Haftungsrecht und Standardbildung in der modernen Medizin: e:med und Probleme der Definition des Standards, MedR 2016, 669 (670 ff.); ausf. M. Keil, Rechtsfragen der individualisierten Medizin, Berlin Heidelberg 2015; A. Ernst, Rechtsfragen der Systemmedizin, Berlin Heidelberg 2020.

6 Informationen unter www.intuitive.com; s. auch www.wikipedia.org/wiki/Da-Vinci-Operationssystem.

auf Grundlage dessen die chirurgischen Instrumente optimal lenken oder sich selbständig neue Operationstechniken erarbeiten.⁷

II. Spezifische Risiken

Trotz aller Fortschritte und Verheißungen besteht gerade in der Medizin verbreitet ein gewisses Unbehagen gegenüber Robotik, die eine Enthumanisierung von Krankenversorgung und Pflege befördern kann, und gegenüber algorithmischer Entscheidungsfindung (ADM).⁸ Aufgrund potentiell undurchsichtiger Entscheidungskriterien werden Befürchtungen geäußert, ihr Einsatz fördere Machtmissbrauch und Manipulation im Gesundheitswesen.⁹ Deutlich werden die durch Digitalisierung ausgelösten Konfliktlinien zwischen Innovation, medizinischem Fortschritt und technischer Machbarkeit einerseits und individuellen Rechtsgütern, objektiven Wertprinzipien sowie ethischen Grundprinzipien andererseits.¹⁰ Es ist ungewiss, wie die medizinethischen Prinzipien Respekt vor der Selbstbestimmung des Patienten, Schadensvermeidung, Fürsorge und Gerechtigkeit¹¹ beim Einsatz von KI und Entscheidungsfindung durch autonome Systeme abgesichert werden.

7 L. Mühlböck/J. Taupitz, Haftung für Schäden durch KI in der Medizin, AcP 221 (2021), 179 (181 f.); H. Lindenberg, Die Zukunft des Robodoktors, ZMGR 2020, 12 (13); s. auch O. Brand, Haftung und Versicherung beim Einsatz von Robotik in Medizin und Pflege, MedR 2019, 943 (947); F. Münch, Autonome Systeme im Krankenhaus, Baden-Baden 2017, S. 52.

8 Vgl. die Studie „Was Europa über Algorithmen weiß und denkt“, Ergebnisse einer repräsentativen Bevölkerungsumfrage im Auftrag der Bertelsmann-Stiftung, Gütersloh 2019, S. 7, 27.

9 Deutscher Ethikrat, Big Data und Gesundheit, Berlin 2017, S. 18 f.: „großes Missbrauchspotenzial“; zu Manipulationen ärztlicher Entscheidungen zugunsten der Produkte eines best. Arzneimittelherstellers J. Taupitz, Medizinische Informationstechnologie, leitliniengerechte Medizin und Haftung des Arztes, AcP 211 (2011), 352 (384) m. Rspr.-Nachw.

10 Katzenmeier, Big Data (Fn. 1), 268 f.; zu rechtlichen Implikationen der Nachvollziehbarkeit von KI-Anwendungen (Explainable AI) in der Medizin s. S. Hänold/N. Schlee/D. Antweiler/K. Beckh, Die Nachvollziehbarkeit von KI-Anwendungen in der Medizin, MedR 2021, 516.

11 Vier-Prinzipien-Modell nach T. Beauchamp/J. Childress, Principles of Biomedical Ethics, 7th ed., Oxford: Oxford University Press 2012; dazu G. Maio, Mittelpunkt Mensch: Ethik in der Medizin, 2. Aufl., Stuttgart 2017, S. 157 ff.; H. Krefß, Medizinische Ethik, 2. Aufl., Stuttgart 2009, S. 15 ff.; zur Geltung als ethische Leitplanken auch in einer digitalisierten Medizin C. Wooten, Vortrag 120, DÄT 2017.

Soweit in der Diskussion von Schadenzufügung durch „*autonome Systeme*“ gesprochen wird, ist nicht die personale Autonomie von Menschen gemeint, also die Fähigkeit zu freiheits- und vernunftmäßigem Handeln im Zeichen moralischer Imperative. Vielmehr geht es um einen Minimalbegriff der *technischen Autonomie*, der speziell auf das Haftungsrecht und seine Funktionen zugeschnitten ist.¹² Danach ist ein System autonom, wenn die Wahl einer bestimmten Verhaltensoption nicht determiniert, sondern das Verhalten des KI-Systems von menschlichen Handlungsanweisungen emanzipiert ist, sodass sich die Entscheidung auch von seinem Programmierer nicht sicher vorhersagen lässt.¹³

Der geringe Einfluss des Menschen auf die konkrete Entscheidung autonomer Systeme wirft die Frage auf, wer für Schädigungen bei deren Einsatz haftbar gemacht werden kann.¹⁴ Das *Autonomierisiko* gilt es rechtlich zu klären, auch wenn aktuell in der Medizin noch kaum autonome Systeme im Einsatz sind, die Schäden unmittelbar verursachen – wie dies im Straßenverkehr bei selbstfahrenden Kfz der Fall ist –, vielmehr in aller Regel Menschen die Entscheidungen der KI umsetzen und die medizinische Behandlung durchführen.¹⁵ Angesichts der Anthropozentrierung unseres Haftungsrechts besteht die Sorge, dass das Recht nicht auf den Einsatz von KI vorbereitet ist und der durch ein autonomes System Geschädigte keinen Ersatz erlangen kann, sodass eine „*Haftungslücke*“ entsteht.¹⁶ Die spezifischen Risiken dieser

12 G. Teubner, Digitale Rechtssubjekte?, AcP 218 (2018), 155 (173); G. Wagner, Verantwortlichkeit im Zeichen digitaler Technik, VersR 2020, 717 (720); G. Spindler, Medizin und IT, insbesondere Arzthaftungs- und IT-Sicherheitsrecht, in: C. Katzenmeier (Hrsg.), Festschrift für Dieter Hart, Berlin Heidelberg 2020, S. 581 (583). Näher zum Autonomiebegriff in Bezug auf Systeme T. Schulz, Verantwortlichkeit bei autonom agierenden Systemen, Baden-Baden 2015, S. 43 ff.

13 S. Russell/P. Norvig, Künstliche Intelligenz, 3. Aufl., München 2012, S. 66; im juristischen Kontext H. Zech, Künstliche Intelligenz und Haftungsfragen, ZfPW 2019, 198 (200); Wagner, Verantwortlichkeit (Fn. 12), 720; Spindler, in: C. Katzenmeier (Hrsg.), FS Hart (Fn. 12), S. 581 (583).

14 Mit zunehmender Autonomie kann sich auch die Frage des Arztvorbehalts stellen, § 1 HPG.

15 Der bis zum Jahr 2004 eingesetzte computergestützte Operationsroboter *Robodoc* (zu Haftungsfragen s. BGHZ 168, 103 = NJW 2006, 2477 m. Bespr. C. Katzenmeier, NJW 2006, 2738) war kein autonomes System und auch bei dem roboter-assistierten Chirurgesystem *Da Vinci* handelt es sich (noch) lediglich um ein automatisiertes System (Berichte über Zwischenfälle unter <https://de.wikipedia.org/wiki/Da-Vinci-Operationssystem>).

16 Teubner, Rechtssubjekte (Fn. 12), 157 ff., 185 f., 189 u. passim; s. auch G. Spindler, Digitale Wirtschaft – analoges Recht: Braucht das BGB ein Update?, JZ 2016, 805 (816); R. Schaub, Interaktion von Mensch und Maschine, JZ 2017, 342 (346); Wagner,

Systeme, ihre *Autonomie*, die *Vernetzung* und die *geringe Transparenz* der Abläufe werfen die Frage nach der Belastbarkeit des geltenden Rechts auf.¹⁷

B. Haftung *de lege lata*

Bislang gibt es keine speziellen Vorschriften zur Haftung für Schäden infolge des Einsatzes von KI. Auch der von der EU-Kommission am 21.4.2021 vorgestellte Entwurf für den *Artificial Intelligence Act*¹⁸ sowie der darauf aufbauende Vorschlag für eine Richtlinie zur Anpassung der Vorschriften über außervertragliche zivilrechtliche Haftung an künstliche Intelligenz vom 28.9.2022¹⁹ enthält keine Haftungsregelung. Autonome Systeme unterliegen damit den allgemeinen Haftungsnormen. Bei der Analyse und Bewertung dürfen menschliches Handeln und Maschinenhandeln nicht unreflektiert gleichgesetzt werden.²⁰ Das autonome System ist nicht lediglich Werkzeug in der Hand des Anwenders.²¹ Es geht darum, *das Fehlverhalten des Systems* in „seiner“ Entscheidungssituation zu erfassen und zu würdigen.²²

Verantwortlichkeit (Fn. 12), 734; Mühlböck/Taupitz, Haftung (Fn. 7), 210; *Datenethikkommission*, Gutachten, Berlin 2019, S. 219 ff.; M. Sommer, Haftung für autonome Systeme, Baden-Baden 2020, passim; M. Thöne, Autonome Systeme und deliktische Haftung, Tübingen 2020, passim.

- 17 Teubner, Rechtssubjekte (Fn. 12), 164; monographisch D. Linardatos, Autonome und vernetzte Aktanten im Zivilrecht, Tübingen 2021; T. Voigt, Produkthaftung, Tübingen 2023 (im Erscheinen).
- 18 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) on 21.4.2021 – COM (2021) 206 final; dazu G. Spindler, Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E), CR 2021, 361; C. Geminn, Die Regulierung Künstlicher Intelligenz, ZD 2021, 354.
- 19 COM (2022) 496 final. Der noch nicht verabschiedete Richtlinienentwurf zielt auf Mindestharmonisierung einzelner Tatbestandsmerkmale des bestehenden nationalen Haftungsrechts, nicht auf die Schaffung neuer Haftungsnormen (s. insb. S. 3 a.E., S. 12 ff. und Erw.-Grd. 14).
- 20 Wagner, Verantwortlichkeit (Fn. 12), 724.
- 21 Kritisch zu dieser noch vorherrschenden Ansicht Teubner, Rechtssubjekte (Fn. 12), 156 ff.
- 22 Teubner, Rechtssubjekte (Fn. 12), 155 f., 170 ff., 192 f., 204; Wagner, Verantwortlichkeit (Fn. 12), 724, s. auch 726: Definitionsgemäß sind autonome Systeme fähig zu autonomem Verhalten. Das wird immer wieder verkannt oder in Abrede gestellt, vgl. etwa die Diskussionsbeiträge auf dem Karlsruher Forum 2020 von A. Bruns, zit. nach N. Biller-Bomhardt/F. Kunz, Karlsruher Forum 2020 über die Verantwortlichkeit im Zeigen digitaler Technik, VersR 2020, 755 (759): eine Maschine kann keine Fehler begehen, sondern nur durch falsche Programmierung fehlerhaft sein; auch P. Pohlmann, zit. nach

I. Haftung des Anwenders

1. Vertragliche Haftung

Der Anwender – in der Medizin Arzt oder Krankenhausträger – haftet bei einem mit ihm wirksam geschlossenen Behandlungsvertrag für Pflichtverletzungen gemäß §§ 630a, 280 Abs. 1 BGB.

a) Pflichten des Arztes/Krankenhausträgers

aa) Den Behandelnden trifft ein strenges Pflichtenprogramm bezüglich der gesamten Behandlung: von der Anamnese über die Untersuchung, Diagnose, Prophylaxe, Therapie bis zur Nachsorge, überdies Organisationspflichten zur Gewährleistung der Patientensicherheit.²³ Beim Einsatz technischer Geräte gelten hohe Anforderungen an Sicherheits- und Kontrollvorkehrungen. Von dem Arzt/Krankenhausträger wird erwartet, dass er stets Apparate verwendet, die dem Erkenntnisstand der medizinischen Wissenschaft entsprechen, dass er diese regelmäßig durch das dafür zuständige Fachpersonal warten lässt, dass er sich im Umgang mit den modernen Techniken schult und fortbildet, dass er die Bedienungsanweisungen genau beachtet und befolgt und dass er das ordnungsgemäße Funktionieren der Apparate fortlaufend überwacht.²⁴

Die Pflicht des Behandelnden, Geräte so einzusetzen, dass Schäden durch ihren Umgang möglichst vermieden werden, intensiviert sich aufgrund der Unvorhersehbarkeit von KI erheblich.²⁵ Der Arzt kann aber keine Garantie

Biller-Bomhardt/Kunz, *Karlsruher Forum* 2020 (Fn. 2), 761; dagegen *G. Wagner*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum* 2020 (Fn. 22), 761.

23 Vgl. allg. *C. Katzenmeier*, in: *A. Laufs/C. Katzenmeier/V. Lipp*, *Arztrecht*, 8. Aufl., München 2021, Kap. X Rn. 3 f. und 41 ff.

24 Auch zu den aus § 3 MPBetreibV folgenden Pflichten des Betreibers eines Medizinprodukts *Spindler*, in: *C. Katzenmeier* (Hrsg.), *FS Hart* (Fn. 12), S. 581 (586 ff.); weitere Nachweise bei *Katzenmeier*, in: *Arztrecht* (Fn. 23), Kap. XI Rn. 126; wichtig der Hinweis von *Hart*, *Haftungsrecht* (Fn. 5), 675: „Haftungsrechtlich wird sich Verantwortlichkeit für intra- und inter-/multidisziplinäre und teletechnologische Zusammenarbeit eher als Organisationshaftung und als Haftung für die Qualität und Sicherheit von Therapieprogrammen darstellen. Behandlungsfehler werden eher zu Programmanwendungsfehlern“.

25 *Spindler*, in: *C. Katzenmeier* (Hrsg.), *FS Hart* (Fn. 12), S. 581 (588); *ders.*, in: *B. Gsell/W. Krüger/S. Lorenz/C. Reymann* (Hrsg.), *Beck-OGK-BGB*, München Stand 1.5.2021, § 823 Rn. 1055; *G. Wagner*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum* 2020 (Fn. 22), 761.

für das fehlerfreie Funktionieren der von ihm eingesetzten medizinisch-technischen Apparate übernehmen. Der Mediziner ist „kein Techniker im Arztkittel“.²⁶ Auch ist der Einsatz eines den genannten Anforderungen genügenden autonomen Systems *nicht per se pflichtwidrig*.²⁷ Stimmen, die dies anders sehen,²⁸ erheben die Autonomie des Systems zum Haftungsgrund des Anwenders. Das aber verträgt sich nicht mit dem geltenden Verschuldensprinzip. Bei einem unerkennbaren Fehler des autonomen Systems liegt keine Sorgfaltspflichtverletzung des Anwenders im Sinne der §§ 630a Abs. 2, 276 BGB vor.

Auch greift die Fehlervermutung des § 630h Abs. 1 BGB nicht, denn dafür reicht es nicht aus, dass der Einsatz des Systems als solcher beherrschbar ist, vielmehr ist die volle Beherrschbarkeit des Systems in der konkreten Situation erforderlich, angesichts dessen Autonomie aber nicht gegeben.²⁹ Nur im Falle eines für den Behandelnden vermeidbaren Versagens der Geräte greift die Vermutung, dass die Pflicht zur Gewährleistung entsprechender Sicherheit nicht genügend beachtet worden ist.³⁰

bb) Neben der Pflicht zu sorgfältiger Behandlung kann die *Aufklärungspflicht* haftungsrelevant werden. Der Behandelnde hat den Patienten über sämtliche für die Einwilligung wesentlichen Umstände aufzuklären, § 630e Abs. 1 S. 1 BGB. Hierzu gehören insbesondere Art, Umfang, Durchführung, zu erwartende Folgen und Risiken der Maßnahme sowie ihre Notwendigkeit, Dringlichkeit, Eignung und Erfolgsaussichten im Hinblick auf die Diagnose oder die Therapie.³¹ Eine Aufklärung des Patienten *durch KI* wäre arztentlastend, ist aber nicht statthaft, § 630e Abs. 2 Nr. 1 BGB.³² Über *Behandlungsalternativen* ist aufzuklären, wenn die alternative Methode entweder bei

26 D. Giesen, *Arzthaftungsrecht im Umbruch*, JZ 1982, 345 (349); Katzenmeier, in: *Arztrecht* (Fn. 23), Kap. XI Rn. 127.

27 Es kann der Verzicht auf den Einsatz verfügbarer KI rechtfertigungsbedürftig sein, nämlich wenn diese höhere Sicherheit bietet als menschliches Handeln, Taupitz, *Informationstechnologie* (Fn. 9), 386; Katzenmeier, *Big Data* (Fn. 1), 268.

28 H. Zech, *Entscheidungen digitaler autonomer Systeme*, Gutachten zum 73. DJT, München 2020, S. A 55; ders., *Zivilrechtliche Haftung für den Einsatz von Robotern – Zuweisung von Automatisierungs- und Autonomierisiken*, in: S. Gless/K. Seelmann (Hrsg.), *Intelligente Agenten und das Recht*, Baden-Baden 2016, S. 163 (191 ff).

29 Spindler, in: C. Katzenmeier (Hrsg.), *FS Hart* (Fn. 12), S. 581 (594); ders. (Fn. 25), § 823 Rn. 1057; a.A. Brand, *Haftung* (Fn. 7), 950.

30 G. Wagner, in: F. J. Säcker/R. Rixecker/H. Oetker/B. Limperg (Hrsg.), *MüKo-BGB*, 8. Aufl., München 2020, § 630h Rn. 26.

31 Näher Katzenmeier, in: *Arztrecht* (Fn. 23), Kap. V Rn. 26 ff.

32 Zum Aufklärungspflichtigen vgl. Katzenmeier, in: *Arztrecht* (Fn. 23), Kap. V Rn. 46 ff.; zur Delegation der Aufklärung C. Katzenmeier/C. Achterfeld, *Delegation der Selbstbes-*

gleichwertiger Heilungs- oder Erfolgsaussicht eine geringere Risikobelastung des Patienten aufweist oder bei nach Art und Richtung gleichwertigen Belastungen und Risiken eine größere Heilungs- oder Erfolgsaussicht verspricht, § 630e Abs. 1 S. 3 BGB.

Will der Arzt vom medizinischen Standard abweichen und eine neue, noch *nicht allgemein eingeführte Methode* mit noch nicht abschließend geklärten Risiken anwenden, muss er den Patienten darüber aufklären, dass es sich um eine neue Methode handelt, die noch nicht lange praktiziert wird, deren Wirksamkeit statistisch noch nicht abgesichert ist und bei der *unbekannte Risiken* nicht auszuschließen sind. Der Patient muss in die Lage versetzt werden, sorgfältig abzuwägen, ob er sich nach der herkömmlichen Methode mit bekannten Risiken behandeln lassen möchte oder nach der neuen Methode unter besonderer Berücksichtigung der in Aussicht gestellten Vorteile und der noch nicht in jeder Hinsicht bekannten Gefahren.³³

Solange der Einsatz von KI eine neue Methode darstellt, ist über diesen Umstand aufzuklären³⁴ und über die damit verbundenen, auch unbekanntes Risiken. Offen ist derzeit, ob und inwieweit nach Etablierung von KI als Standardmethode über das weiterhin bestehende *Autonomierisiko* aufzuklären ist. Soweit es sich dabei um einen für die Entscheidungsfreiheit und damit das Selbstbestimmungsrecht des Patienten relevanten Umstand handelt, ist er darüber in Kenntnis zu setzen.

b) Digitale Erfüllungsgehilfen?

Im Schrifttum wird vertreten, autonome Systeme seien als *digitale Erfüllungsgehilfen* zu qualifizieren und ihr Fehlverhalten dem Vertragsschuldner mit haftungsbegründender Wirkung analog § 278 S. 1 BGB zuzurechnen.³⁵

timungsaufklärung, in: H. Steinmeier/N. Roeder/W. v. Eiff (Hrsg.), Festschrift für Karl Otto Bergmann, Berlin Heidelberg 2016, S. 89 ff.

33 BGHZ 168, 103 (109) (Robodoc) = NJW 2006, 2477 (2478 f.) m. Anm. C. Katzenmeier NJW 2006, 2738 (2740); BGHZ 172, 254 (260 ff.) (Racz-Katheter) = NJW 2007, 2774 (2775) = MedR 2008, 87 (88 f.) m. Anm. A. Spickhoff; zur neueren BGH-Rspr. M. Vogeler, Die Haftung des Arztes bei der Anwendung neuartiger und umstrittener Heilmethoden nach der neuen Rechtsprechung des BGH, MedR 2008, 697 (704 ff.); s. auch BGH NJW 2020, 1358, 1360 = MedR 2020, 379, 381 m. Anm. C. Jansen.

34 Spindler, in: C. Katzenmeier (Hrsg.), FS Hart (Fn. 12), S. 581 (592); s. auch S. Beck, Zum Einsatz von Robotern im Palliativ- und Hospizbereich, MedR 2018, 772 (776).

35 Teubner, Rechtssubjekte (Fn. 12), 186 ff.; J.E. Schirmer, Rechtsfähige Roboter?, JZ 2016, 660 (664 f.); bei vollständig autonomen Systemen P. Hacker, Verhaltens- und Wissens-

Dahinter steht der Gedanke, dass Anwender, die digitale Systeme einsetzen, nicht privilegiert werden sollen gegenüber denen, die Menschen beschäftigen.³⁶ Durch eine analoge Anwendung des § 278 S. 1 BGB ließe sich die gleiche Risikoverteilung wie beim Einsatz menschlicher Hilfspersonen erzielen.

Dagegen wird eingewandt, dass es sich bei KI-Systemen trotz ihrer Künstlichen Intelligenz und Autonomie um nicht rechtsfähige technische Einrichtungen handelt.³⁷ Mangels Rechtsfähigkeit können KI-Systeme nicht Träger von Rechten und Pflichten und somit auch nicht Normadressat sein. Fehle es an einem Zurechnungssubjekt, funktioniert die Zurechnungsmechanik des § 278 S. 1 BGB nicht und die Analogie scheitert.³⁸ Erkennt man in dieser Argumentation einen Zirkelschluss,³⁹ so ist doch maßgeblich, dass autonomen Systemen *kein Verschuldensvorwurf* gemacht werden kann, da ein solcher trotz Objektvierung der Fahrlässigkeit immer auch eine subjektive Komponente umfasst.⁴⁰ Der Arzt haftet nach Vertragsrecht also für eigene Pflichtverletzungen beim Einsatz von KI, nicht aber wird ihm ein Fehlverhalten des autonomen Systems zugerechnet.

2. Deliktische Haftung

a) Verkehrspflichten

Nach § 823 Abs. 1 BGB haftet, wer rechtswidrig und schuldhaft ein Rechtsgut eines anderen verletzt. Die Sorgfaltsgebote des Deliktsrechts sind verbunden mit dem Begriff der Verkehrspflichten, andere nicht mehr als unvermeidbar zu gefährden.⁴¹ Inhaltlich sind die Verkehrspflichten bei der Arzthaftung weitestgehend deckungsgleich mit den vertraglichen Behandlungs- und Aufklärungspflichten, die hier ebenfalls ganz auf den Schutz von Körper, Ge-

zurechnung beim Einsatz von Künstlicher Intelligenz, RW 2018, 243 (251 ff.); bzgl. Maschinen und Automaten s. bereits K. *Spiro*, Die Haftung für Erfüllungsgehilfen, 1984, S. 211; M. *Wolf*, Schuldnerhaftung bei Automatenversagen, JuS 1989, 899 (901); vgl. auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss Nr. 8.

36 *Teubner*, Rechtssubjekte (Fn. 12), 188; *Zech*, Entscheidungen (Fn. 28), S. A 76.

37 S. *Klingbeil*, Schuldnerhaftung für Roboterversagen, JZ 2019, 718 (719); *Spindler*, in: C. *Katzenmeier* (Hrsg.), FS Hart (Fn. 12), S. 581 (585).

38 *Brand*, Haftung (Fn. 7), 950; *Mühlböck/Taupitz*, Haftung (Fn. 7), 198.

39 C. *Linke*, Die elektronische Person, MMR 2021, 200 (203).

40 *Schaub*, Interaktion (Fn. 16), 343; *Mühlböck/Taupitz*, Haftung (Fn. 7), 198, 200.

41 C. v. *Bar*, Verkehrspflichten, Köln 1980; C. *Katzenmeier*, in: B. *Dauner-Lieb/W. Langen* (Hrsg.), NK-BGB, Bd. 2, 4. Aufl., Baden-Baden 2021, § 823 Rn. 124 ff. m. w. N.

sundheit und Selbstbestimmungsrecht des Patienten bezogen sind.⁴² Wie bei der vertraglichen gilt auch bei der deliktischen Haftung, dass der Einsatz eines autonomen Systems *nicht per se pflichtwidrig* ist, also auch keine Haftung gemäß § 823 Abs. 1 BGB begründet.

b) Maschinelle Verrichtungsgehilfen

In Betracht zu ziehen ist aber eine Haftung analog § 831 BGB. Nach § 831 Abs. 1 S. 1 BGB ist derjenige, der einen anderen zu einer Verrichtung bestellt, zum Ersatz des Schadens verpflichtet, den der andere einem Dritten widerrechtlich zufügt. Voraussetzung ist lediglich, dass der Gehilfe tatbestandsmäßig und rechtswidrig im Sinne der §§ 823 ff. BGB gehandelt hat. Auf ein Verschulden des Verrichtungsgehilfen kommt es nach dem Wortlaut der Norm nicht an.⁴³

Autonomen Systemen wird aufgrund ihrer derzeitigen kognitiven Fähigkeiten von Seiten der Informationsphilosophie, die die moralische Verantwortlichkeit von Algorithmen untersucht, zwar (noch) keine *responsability* im Vollsinn zugesprochen, jedoch eine *accountability* – was in die juristische Dogmatik übersetzt bedeutet, dass sie zwar nicht schuldhaft im Sinne subjektiver Vorwerfbarkeit, immerhin aber doch rechtswidrig handeln können.⁴⁴ Sie lassen sich daher als *maschinelle Verrichtungsgehilfen* qualifizieren.⁴⁵

Ausgestaltet ist die Geschäftsherrenhaftung als eine Haftung für vermutetes eigenes Auswahl- und Überwachungsverschulden, was dem anspruchstellenden Patienten günstig sein kann. Allerdings erweist sich die in § 831

42 Katzenmeier, in: *Arztrecht* (Fn. 23), Kap X Rn. 2 u. XI Rn. 63.

43 Vorbild der Haftung nach § 831 BGB ist auch die römisch-rechtliche Noxalhaftung des *pater familias* für seinen Sklaven, dem es – wie dem autonomen System – an Rechtsfähigkeit fehlte, *F. Bernau*, in: J. v. Staudinger (Begr.), BGB, Buch 2, Berlin 2018, § 831 Rn. 1.

44 *Teubner*, *Rechtssubjekte* (Fn. 12), 188 (für eine Qualifikation autonomer Systeme als digitale Erfüllungsgehilfen) mit Verweis auf *L. Floridi*, in: M. Anderson/S. Anderson, *Machine Ethics*, Cambridge, Mass.: Cambridge University Press 2011, S. 184, 205 ff.; *Mühlböck/Taupitz*, *Haftung* (Fn. 7), 199; a.A. *Schaub*, *Interaktion* (Fn. 16), 344; *Brand*, *Haftung* (Fn. 7), 949.

45 *T. Riehm*, *Von Drohnen, Google-Cars und Software-Agenten*, ITRB 2014, 113 (114); *M. Denga*, *Deliktische Haftung für künstliche Intelligenz*, CR 2018, 69 (74 ff.); *Hacker*, *Verhaltens- und Wissenszurechnung* (Fn. 35), 265 ff.; *Spindler*, in: C. Katzenmeier (Hrsg.), *FS Hart* (Fn. 12), S. 581 (596); so auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschlüsse Nr. 9–11; mangels Rechtssubjektivität von KI zweifelnd *Voigt*, *Produkthaftung* (Fn. 17), § 5 C. IV. 1. b).

Abs. 1 S. 2 BGB vorgesehene *Exkulpationsmöglichkeit* des Geschäftsherrn gerade beim Einsatz autonomer Systeme als Schwachstelle. Da der Anwender das Verhalten des autonomen digitalen Systems nicht beeinflussen kann, wird er sich zumeist darauf berufen können, dass er seine Pflichten erfüllt hat (oder dass der Schaden unabhängig davon eingetreten ist). Der Entlastungsbeweis wird ihm also regelmäßig gelingen.⁴⁶

c) Digitale Tiere?

Einzelne Stimmen im Schrifttum wollen die Gefährdungshaftung des § 833 S. 1 BGB auf die „*digitale Gefahr*“ erstrecken, der Zurechnungsgrund der Tierhalterhaftung treffe auch auf den Einsatz Künstlicher Intelligenz zu.⁴⁷ Doch ist eine Analogie angesichts des im deutschen Recht geltenden Enumerationsprinzips nicht statthaft. Eine Gefährdungshaftung gilt als Ausnahme vom Grundsatz der Verschuldenshaftung nur dort, wo der Gesetzgeber sie ausdrücklich angeordnet hat.⁴⁸

Abgesehen davon sind autonome Systeme nicht wie digitale Tiere zu behandeln. Das spricht auch gegen eine an sich mögliche Analogie zu der Haftung für vermutetes Verschulden für Nutztiere gemäß § 833 S. 2 BGB. Autonome Systeme sind wertungsmäßig Verrichtungsgehilfen gleichgestellt, die im Interesse eines anderen tätig werden, an dessen Weisungen gebunden sind, im Rahmen eines bestimmten Spielraums jedoch selbst Entscheidungen treffen. § 833 BGB ist dagegen auf Lebewesen zugeschnitten, die zu rationalem Entscheiden zwischen Verhaltensoptionen nicht in der Lage sind, sondern instinktiv reagieren.⁴⁹

46 Hacker, Verhaltens- und Wissenszurechnung (Fn. 35), 266; K. Helle, Intelligente Medizinprodukte: Ist der geltende Rechtsrahmen noch aktuell?, MedR 2020, 993 (998); Wagner, Verantwortlichkeit (Fn. 12), 730; Mühlböck/Taupitz, Haftung (Fn. 7), 204; s. aber auch Spindler, in: C. Katzenmeier (Hrsg.), FS Hart (Fn. 12), S. 581 (596).

47 S. Horner/M. Kaulartz, Haftung 4.0, InTer 2016, 22 (24); G. Borges, Rechtliche Rahmenbedingungen für autonome Systeme, NJW 2018, 977 (981); T. Riehm/S. Meier, Künstliche Intelligenz im Zivilrecht, in: V. Fischer/P. Hoppen/J. Wimmers (Hrsg.), DGRI Jahrbuch, Köln 2018, Rn. 25.

48 BGHZ 55, 229, 234 = NJW 1971, 607 (608); 63, 234 (237) = NJW 1975, 117 (118); K. Lorenz/C. Canaris, SchuldR Bd. II/2, 13. Aufl., München 1994, § 84 I b); D. Looschelders, SchuldR BT, 17. Aufl., München 2022, § 73 Rn. 5 f.

49 Wagner, Verantwortlichkeit (Fn. 12), 731; abw. Brand, Haftung (Fn. 7), 949, der für einen § 833a BGB plädiert.

II. Haftung des Herstellers

Neben dem Anwender kommt insbesondere der Hersteller eines autonomen Systems als Haftungsschuldner in Betracht.⁵⁰ Autonome Systeme können Medizinprodukte sein. Dann finden die einschlägigen Regelwerke, namentlich die Medizinprodukte-Verordnung⁵¹ und das Medizinprodukte-Durchführungsgesetz⁵² Anwendung. Doch enthält weder die MDR noch das MPDG spezielle Haftungsregeln für den Hersteller, sodass auf das Produkthaftungsgesetz und die deliktische Produzentenhaftung nach § 823 Abs. 1 BGB zurückzugreifen ist.⁵³

1. Produkthaftung

Nach § 1 Abs. 1 S. 1 ProdHaftG ist der Hersteller eines Produkts zum Schadensersatz verpflichtet, wenn jemand durch den Fehler des Produkts an seinen Rechtsgütern verletzt wird. Ein Verschulden ist nicht erforderlich.

Produkt ist nach § 2 ProdHaftG jede bewegliche Sache. Ob Daten, online überspielte Software und andere rein digitale Güter darunterfallen, ist streitig. Nach teilweise vertretener Ansicht verkörpern sie lediglich eine geistige Leistung.⁵⁴ Der Schutzzweck des Gesetzes spricht aber für eine Einbeziehung.⁵⁵

50 Weitere Haftungsschuldner können etwa Entwickler eines autonomen Systems, Programmierer, Datenlieferanten, Netzwerkbetreiber, Systemadministratoren sein, vgl. *Spindler*, in: C. Katzenmeier (Hrsg.), FS Hart (Fn. 12), S. 581 (594).

51 Verordnung (EU) 2017/745 über Medizinprodukte (Medical Device Regulation (MDR)) vom 25.5.2017, seit dem 26.5.2021 verpflichtend anzuwenden.

52 Gesetz zur Durchführung unionsrechtlicher Vorschriften betreffend Medizinprodukte (Medizinprodukte-Durchführungsgesetz (MPDG)) vom 28.4.2020, BGBl. I 960, Inkrafttreten überwiegend am 26.5.2021.

53 *I. Jakobs/F. Huber*, Software als Medizinprodukt: Haftungs- und versicherungsrechtliche Aspekte, MPR 2019, 1; *Y. Frost/M. Kießling*, Künstliche Intelligenz im Bereich des Gesundheitswesens und damit verbundene haftungsrechtliche Herausforderungen, MPR 2020, 178 (179 ff.); *Helle*, Medizinprodukte (Fn. 46), 996; *Mühlböck/Taupitz*, Haftung (Fn. 7), 187.

54 *J. Oechsler*, in: J. v. Staudinger (Begr.), BGB, Buch 2, Berlin 2021, § 2 ProdHaftG Rn. 65 m.w.N.; Produkteigenschaft auch zu verneinen nach EuGH, NJW 2021, 2015 m. Anm. *M. Finkelmeier*.

55 *G. Spindler*, IT-Sicherheit und Produkthaftung – Sicherheitslücken, Pflichten der Hersteller und der Softwarenutzer, NJW 2004, 3145 (3149); *G. Wagner*, Produkthaftung für autonome Systeme, AcP 217 (2017), 707 (716 ff.); *P. Hacker*, Europäische und nationale Regulierung von Künstlicher Intelligenz, NJW 2020, 2142 (2145); *Wagner* (Fn. 30), § 2 ProdHaftG Rn. 18 f.; *Katzenmeier* (Fn. 41), § 2 ProdHaftG Rn. 3; *C. Katzenmeier/T. Vo-*

EU-Kommission und Bundesregierung haben eine normative Klarstellung angekündigt, dass auch nicht integrierte Software in den Anwendungsbereich der Produkthaftung fallen soll.⁵⁶

Nach § 3 ProdHaftG hat ein Produkt einen Fehler, wenn es nicht die Sicherheit bietet, die unter Berücksichtigung aller Umstände berechtigterweise erwartet werden kann. Bei der Bestimmung der für ein KI-System geltenden Sicherheitsanforderungen stellt sich die Frage nach dem Vergleichsmaßstab. Man kann das Inverkehrbringen als pflichtwidrig qualifizieren und einen Fehler bejahen, wenn das KI-System nicht dieselbe Sicherheit bietet wie ein ähnliches von Menschen gesteuertes System. Im Übrigen aber ist es aufgrund der technischen Gegebenheiten kaum möglich, ein autonomes System mit dem sorgfältigen Verhalten eines Menschen zu vergleichen.⁵⁷ Mit der Implementierung des lernfähigen Steuerungsalgorithmus in das KI-System wird lediglich dessen grundlegendes Verhalten programmiert, nicht aber die Entscheidung in einer konkreten Situation. Alternativ zum anthropozentrischen ist ein *systembezogener Sorgfaltsmaßstab* zu erwägen, wonach das schädigende System an der Leistung anderer vergleichbarer KI-Systeme zu messen ist.⁵⁸ Aufgabe der Rechtsprechung ist es, im Einzelfall konkrete Pflichtenstandards zu definieren,⁵⁹ wobei die modernen Technologien eine Standardbildung erschweren können.⁶⁰

Große Schwierigkeiten bereitet dem Geschädigten die Beweislast hinsichtlich des Fehlers und der Kausalität, beides festgelegt in § 1 Abs. 4 S. 1 Prod-

igt, ProdHaftG, 7. Aufl., Berlin 2020, § 2 Rn. 16 ff.; Voigt, Produkthaftung (Fn. 17), § 5 B. II.

56 EU-Kommission, Bericht COM (2020) 64, v. 19.2.2020, S. 17; konkreter der Richtlinienvorschlag v. 28.9.2022 – COM (2022) 495, S. 1, 9, 14 sowie Erw.-Grd. 12 zu Art. 4 Abs. 1, S. 19, 29; Bundesregierung, Stellungnahme zum Weißbuch v. 29.6.2020, S. 24 f.; vgl. auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss Nr. 4.

57 Wagner, Produkthaftung (Fn. 55), 734; Mühlböck/Taupitz, Haftung (Fn. 7), 190 f.

58 So auch *Expert Group on Liability and New Technologies*, Liability for Artificial Intelligence and other emerging digital technologies, Luxemburg 2019, S. 46; näher zum systembezogenen Fehlerbegriff Wagner, Produkthaftung (Fn. 55), 735 ff.; Wagner, Verantwortlichkeit (Fn. 12), 727 f.; Mühlböck/Taupitz, Haftung (Fn. 7), 190 f.; im Lichte der jüngeren Rspr. einen weitergehend objektivierten, weniger verhaltens- und sorgfaltszentrierten Fehlerbegriff befürwortend hingegen Voigt, Produkthaftung (Fn. 17), § 5 B. III. u. IV.

59 Teubner, Rechtssubjekte (Fn. 12), 194; zur Standardbildung D. Hart, Evidenz-basierte Medizin und Gesundheitsrecht, MedR 2000, 1; ders., Kongruenz und Kontinuität in der Entwicklung von Medizin und Medizinrecht, MedR 2015, 1; C. Jansen, Der Medizinische Standard, 2019, S. 183 ff.

60 Hart, Haftungsrecht (Fn. 5), 669 ff.

HaftG. Während bei herkömmlichen Gütern der Mangel äußerlich erkennbar oder für einen Sachverständigen feststellbar ist, gestaltet sich die Feststellung eines Fehlers bei digitalen Produkten – bei KI: fehlerhafter Steuerungsalgorithmus, fehlerhaft designtes künstliches neuronales Netz, unvollständiger Trainingsdatensatz – wesentlich aufwendiger.⁶¹ Hinzu kommt, dass es sich bei KI-Systemen nicht um diskrete Systeme handelt, sondern dass sie in vielfältiger Weise mit anderen KI-Systemen oder Datenquellen vernetzt werden. Die *Vernetzung* bietet Einfallstore für Risiken aus fremden Sphären. Das wiederum führt zu komplexeren Kausalverläufen und bereitet Schwierigkeiten bei der Zurechnung und Aufklärung von Geschehensabläufen.⁶²

Schließlich ist auch die verschuldensunabhängige Haftung des Herstellers gemäß § 1 Abs. 2 Nr. 2 ProdHaftG ausgeschlossen, wenn das Produkt bei Inverkehrbringen fehlerfrei war, überdies gemäß § 1 Abs. 2 Nr. 5 ProdHaftG, wenn der Fehler bei Inverkehrbringen nicht erkennbar war. Damit werden gerade diejenigen Risiken als von der Haftung ausgenommen erachtet, die aus unvorhersehbarem Verhalten eines autonomen Systems resultieren.⁶³

2. Produzentenhaftung

Während das Produkthaftungsgesetz auf den Zeitpunkt des Inverkehrbringens abstellt, enden die Verkehrspflichten des Herstellers nach § 823 Abs. 1

61 P. Reusch, in: M Kaulartz/T. Braegelmann (Hrsg.), *Rechtshandbuch Artificial Intelligence und Machine Learning*, 2020, S. 124 ff.; G. Borges, *Haftung für selbstfahrende Autos*, CR 2016, 272 (275); Schaub, *Interaktion* (Fn. 16), 344; W. Droste, *Intelligente Medizinprodukte: Verantwortlichkeiten des Herstellers und ärztliche Sorgfaltspflichten*, MPR 2018, 109 (113); Wagner, *Verantwortlichkeit* (Fn. 12), 729; Katzenmeier/Voigt, *Produkthaftung* (Fn. 55), Einl. Rn. 24.

62 Zech, *Künstliche Intelligenz* (Fn. 13), 203, 205; s. auch Teubner, *Rechtssubjekte* (Fn. 12), 201; Wagner, *Verantwortlichkeit* (Fn. 12), 734; Spindler, in: C. Katzenmeier (Hrsg.), *FS Hart* (Fn. 12), S. 581 (584, 597); Mühlböck/Taupitz, *Haftung* (Fn. 7), 183 f.; dazu unter III. 1. b.

63 J. Taeger, *Die Entwicklung des IT-Rechts im Jahr 2016*, NJW 2016, 3764 (3765); C. Gomme, *Herstellerhaftung für automatisierte Fahrzeuge*, JZ 2016, 76 (78 f.); A. Seehafer/J. Kohler, *Künstliche Intelligenz: Updates für das Produkthaftungsrecht?*, EuZW 2020, 213 (215); Spindler, in: C. Katzenmeier (Hrsg.), *FS Hart* (Fn. 12), S. 581 (601); Oechsler (Fn. 54), § 3 ProdHaftG Rn. 85a; krit. ggü. der Einordnung des Autonomierisikos als Entwicklungsrisiko Teubner, *Rechtssubjekte* (Fn. 12), 190; programmierte (!) Nichtprognostizierbarkeit (!) der Algorithmus-Entscheidungen; Zech, *Künstliche Intelligenz* (Fn. 13), 213; Konstruktionsfehler; Wagner (Fn. 30), § 1 ProdHaftG Rn. 61; Mühlböck/Taupitz, *Haftung* (Fn. 7), 192 f.; diff. Frost/Kießling, *Künstliche Intelligenz* (Fn. 53), 181 f.

BGB nicht mit dem Inverkehrbringen.⁶⁴ Vielmehr muss er seine Produkte auch danach daraufhin beobachten, welche – bis zu diesem Zeitpunkt unerkannten – Gefahren sie mit sich bringen. Gegebenenfalls muss er eine Warnung aussprechen oder sie zurückrufen. Bei autonomen Systemen verlangt eine zuverlässige *Produktbeobachtung* nicht nur die Auswertung öffentlich zugänglicher Quellen, sondern auch eine Überprüfung der Lernmechanismen und – sofern verfügbar und datenschutzrechtlich zulässig – der zugrundeliegenden Daten und Quellen.⁶⁵ Die Produktbeobachtungspflicht erstreckt sich auch auf Gefahren, die sich erst aus einer Kombination mit Produkten anderer Hersteller ergeben.⁶⁶

Hinsichtlich des bei § 823 Abs. 1 BGB erforderlichen Verschuldens kehrt der BGH die Beweislast zugunsten des Produktgeschädigten um.⁶⁷ Jedoch wird sich der Hersteller bei einem unvorhersehbaren und unvermeidbaren Fehlverhalten des autonomen Systems entlasten können.

3. „Haftungslücke“

Zwischenfazit: Bei Schäden durch den Einsatz autonomer Systeme in der Medizin scheitern vertragliche wie deliktische Ansprüche gegen den anwendenden Arzt oder Krankenhausträger regelmäßig am fehlenden Verschulden. Ansprüche gegen den Hersteller des Systems nach ProdHaftG sind verschuldensunabhängig, aber nur schwer durchsetzbar, da der Geschädigte dazu das Vorliegen eines Produktfehlers und der Kausalität darlegen und im Bestreitensfall beweisen muss, zudem Haftungsausschlussgründe bestehen. Und so konstatieren die Experten eine „*Haftungslücke*“.⁶⁸

Der Begriff Haftungslücke ist freilich irreführend, indem er suggeriert, in jedem Falle eines Schadens müsse es einen Ersatzanspruch geben. Auch wenn

64 Die Haftung bleibt nach § 15 Abs. 2 ProdHaftG unberührt.

65 M. Martini, Algorithmen als Herausforderung für die Rechtsordnung, JZ 2017, 1017 (1021); Droste, Intelligente Medizinprodukte (Fn. 61), 111; Mühlböck/Taupitz, Haftung (Fn. 7), 195.

66 BGHZ 99, 167, 174 – Honda = NJW 1987, 1009 (1010 f.); dazu H. Kullmann, Die Produktbeobachtungspflicht des Kraftfahrzeugherstellers im Hinblick auf Zubehör, BB 1987, 1957; krit. P. Ulmer, Produktbeobachtungs-, Prüfungs- und Warnpflichten eines Warenherstellers in Bezug auf Fremdprodukte?, ZHR 152 (1988), 564 (575 f.); Spindler (Fn. 25), § 823 Rn. 665.

67 GrdI. BGHZ 51, 91 – Hühnerpest = NJW 1969, 269 m. Anm. U. Diederichsen = VersR 1969, 155; dazu C. Katzenmeier, in: G. Baumgärtel/ H. Laumen/H. Prütting (Hrsg.), Handbuch der Beweislast, 5. Aufl., Köln 2022, Bd. 3, § 823 Anhang III Rn. 9 ff.

68 S. sub I. 2.

diese Vorstellung heute im Publikum weit verbreitet ist, lautet der Grundsatz *casum sentit dominus*.⁶⁹ Zweck des Haftungsrechts ist nicht der Ausgleich von Schäden, vielmehr die Bestimmung der Voraussetzungen, unter denen Kompensation zu leisten ist.⁷⁰ Die Überwälzung des Schadens auf einen anderen erfolgt nur, wenn er diesem zugerechnet werden kann. Gründe der Schadenszurechnung sind die rechtswidrig schuldhaft Verursachung oder die Verantwortung für eine besondere Gefahr.⁷¹

4. Haftung de lege ferenda

Die neuartigen digitalen Risiken den Geschädigten als allgemeines Lebensrisiko aufzubürden, wird als „rechtspolitisch verfehlt und fundamental ungerecht“ kritisiert⁷² und eine „genuine Rehumanisierung der technischen Welt“ gefordert.⁷³ Wenn die Gesellschaft neuartige Entscheidungsräume für bis dahin unbekannte autonome Entscheidungsträger zulasse, sei sie verpflichtet, für wirksame Verantwortungsformen zu sorgen.⁷⁴ Dementsprechend hat das Europäische Parlament die Kommission bereits im Jahr 2017 aufgefordert, über neue Haftungsregeln bezüglich Schäden durch den Einsatz von KI nachzudenken.⁷⁵ Die Kommission hat eine Expertengruppe eingesetzt und im Jahr 2020 ein Weißbuch zur Regulierung von KI vorgelegt, das auch

69 C. Waldkirch, *Zufall und Zurechnung im Haftungsrecht*, 2018, S. 201 ff.; M. Rohe, *Gründe und Grenzen deliktischer Haftung – die Ordnungsaufgaben des Deliktsrechts (einschließlich der Haftung ohne Verschulden) in rechtsvergleichender Betrachtung*, AcP 201 (2001), 117 (163 f.); G. Schiemann, *Wandlungen der Berufshaftung*, in: H. Lange/K. Nörr/H. Westermann (Hrsg.), *Festschrift für Joachim Gernhuber*, 1993, S. 387 (392); s. auch H. Mansel, *Eigen- und Fremdverantwortung im Haftungsrecht – zu den außerjuristischen Triebkräften der Haftungsverschärfung*, in: P. Gottwald, *Festschrift für Dieter Henrich* (Hrsg.), Bielefeld 2000, S. 425 (442).

70 *Larenz/Canaris*, *SchuldR II* (Fn. 48), § 75 I. 2. i); *Wagner* (Fn. 30), Vor § 823 Rn. 43 f.; *Katzenmeier* (Fn. 41), Vorbem. zu §§ 823 ff. Rn. 54 f.; L. Loacker, *Abschied vom Ausgleich? Ein Kontrapunkt*, HAVE 2021, 170.

71 W. Canaris, *Grundstrukturen des deutschen Deliktsrechts*, *VersR* 2005, 577; *Wagner* (Fn. 30), Vor § 823 Rn. 43 f.; *Katzenmeier* (Fn. 41), Vorbem. zu §§ 823 ff. Rn. 18 ff., 22 ff.; *Voigt*, *Produkthaftung* (Fn. 17), § 6 E.

72 Vgl. etwa *Teubner*, *Rechtssubjekte* (Fn. 12), 160: „unbarmherziges casum sentit dominus“.

73 *Teubner*, *Rechtssubjekte* (Fn. 12), 176 Fn. 90, mit beißender Kritik an der Dogmatik des geltenden Haftungsrechts.

74 *Teubner*, *Rechtssubjekte* (Fn. 12), 187 f.; sachlich nüchtern *Wagner*, *Verantwortlichkeit* (Fn. 12), 717 f.; *Mühlböck/Taupitz*, *Haftung* (Fn. 7), 186 f.

75 *EU-Parlament*, *Entschließung v. 16.2.2017 – 2015/2103 (INL)*.

Überlegungen zur Fortentwicklung des Haftungsrechts enthält.⁷⁶ In dem am 21.4.2021 veröffentlichten Ersten Entwurf des *Artificial Intelligence Act*⁷⁷ findet sich noch keine konkrete Regelung, doch wird eine solche angekündigt. Auch zwei Richtlinienvorschläge der Kommission vom 28.9.2022 für eine Revision der Produkthaftungsrichtlinie⁷⁸ sowie eine allgemeine Richtlinie zur Anpassung der Vorschriften über außervertragliche zivilrechtliche Haftung an künstliche Intelligenz⁷⁹ zielen nicht auf die Schaffung neuer Haftungsnormen. Nachdem indes zuvor vom Europäischen Parlament die Initiative für eine Verordnung über Haftung für den Betrieb von Systemen mit Künstlicher Intelligenz⁸⁰ ausgegangen war, bleibt die hochaktuelle Thematik umstritten und der Ausgang legislativer Bemühungen offen.

III. Haftungskonzept

1. Gefährdungshaftung

Im Mittelpunkt der Diskussion steht die Einführung einer Gefährdungshaftung, also einer Haftung, die keine Rechtswidrigkeit und kein Verschulden erfordert, sondern lediglich eine Schädigung durch eine besondere, an sich erlaubte Gefahr.⁸¹ Das geltende Produkthaftungsgesetz normiert entgegen

76 *EU-Kommission*, Weißbuch zur Künstlichen Intelligenz v. 19.2.2020 – COM (2020) 65 final, S. 14 ff.; *EU-Kommission*, Bericht zum Weißbuch v. 19.2.2020 – COM (2020) 64 final; *EU-Parlament*, Entschließung v. 20.10.2020 – 2020/2014 (INL).

77 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) on 21.4.2021 – COM (2021) 206 final.

78 COM (2022) 495 final; s. im Vorfeld dazu auch C. Wendehorst et al., *ELI Draft of a Revised Product Liability Directive*, Wien 2022; C. Twigg-Flesner et al., *Guiding Principles for Updating the Product Liability Directive for the Digital Age*, Wien 2019.

79 COM (2022) 496 final.

80 *EU-Parlament*, Entschließung v. 20.10.2020 – 2020/2014 (INL); näher – auch zum legislatorischen Hintergrund – G. Wagner, *Haftung für Künstliche Intelligenz – Eine Gesetzesinitiative des Europäischen Parlaments*, ZEuP 2021, 545.

81 Befürwortend 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss Nr. 6; Schulz, *Verantwortlichkeit bei autonom agierenden Systemen* (Fn. 12), S. 364 ff.; R. Janal, *Die deliktische Haftung beim Einsatz von Robotern – Lehren aus der Haftung für Sachen und Gehilfen*, in: S. Gless/K. Seelmann (Hrsg.), *Intelligente Agenten* (Fn. 28), S. 139 (155); Schirmer, *Roboter* (Fn. 35), 665; Wagner, *Verantwortlichkeit* (Fn. 12), 734 f.; Zech, *Entscheidungen* (Fn. 28), S. A 98 ff.; sektorspezifisch G. Spindler, *User Liability and Strict Liability in the Internet of Things and for Robots*, in: S. Lohsse/R. Schulze/D. Staudenmayer (Hrsg.), *Liability for Artificial Intelligence and the Internet of Things*, Baden-

vorherrschender Ansicht⁸² keine Gefährdungshaftung. Zwar glaubten die Verfasser der dem Gesetz zugrundeliegenden Produkthaftungsrichtlinie ein Regime strikter Haftung zu errichten, tatsächlich handelt es sich aber um eine verschuldensunabhängige Haftung für objektives Verhaltensunrecht.⁸³ Die Haftung knüpft nicht an das Inverkehrbringen eines Produkts, vielmehr an das Inverkehrbringen eines *fehlerhaften* Produkts.⁸⁴

Alle Probleme bezüglich des Fehlernachweises wären gelöst, wenn man das Produkthaftungsrecht vom Fehler abkoppelte und zu einer strikten Haftung weiterentwickelte, die lediglich davon abhängt, dass sich ein spezifisches Risiko realisiert hat, hier die „*digitale Gefahr*“.⁸⁵ Wegen der vollharmonisierenden Wirkung der Produkthaftungs-Richtlinie⁸⁶ könnte das nur auf europäischer Ebene erfolgen. Für eine Gefährdungshaftung treten insbesondere Verfechter der *economic analysis* ein.⁸⁷ Sie sehen in der Präventivfunktion einen

Baden 2019, S. 125 (136 ff.); s. auch *EU-Parlament*, Entschließung v. 20.10.2020 – 2020/2014 (INL), Verordnungsvorschlag, Art. 4 VO-E: Gefährdungshaftung für Betreiber von KI-Systemen mit hohem Risiko.

82 *Larenz/Canaris*, SchuldR II (Fn. 48), § 84 VI. 1.; *H. Taschner*, Die künftige Produzentenhaftung in Deutschland, NJW 1986, 611 (612); *P. Marburger*, Grundsatzfragen des Haftungsrechts unter dem Einfluß der gesetzlichen Regelungen zur Produzenten- und zur Umwelthaftung, AcP 192 (1992), 1 (10 ff.); *W. Rolland*, in: *W. Rolland* (Hrsg.), Produkthaftungsrecht, München 1990, § 1 Rn. 7.

83 *F. G. v. Westphalen*, in: *U. Foerste/F. G. v. Westphalen* (Hrsg.), Produkthaftungshandbuch, 3. Aufl., München 2012, § 45 Rn. 8 f.; *C. v. Bar*, Neues Haftungsrecht durch Europäisches Gemeinschaftsrecht, in: *D. Medicus/H.-J. Mertens/K. Nörr/W. Zöllner* (Hrsg.), Festschrift für Hermann Lange, 1992, S. 373 (390); *Wagner*, Verantwortlichkeit (Fn. 12), 726; *Katzenmeier/Voigt* (Fn. 55), Einl. Rn. 26; *Katzenmeier* (Fn. 41), § 1 ProdHaftG Rn. 1; monographisch *Voigt*, Produkthaftung (Fn. 17), § 3 A. IV. 2 f. und passim, zu jüngeren Rechtsprechungstendenzen einer extensiven Auslegung des Fehlerbegriffs in Annäherung an eine Gefährdungshaftung § 3 A. I. 2. b), § 4 A. III. 2. c), § 5 B. III.–IV.

84 So auch der Richtlinienvorschlag zur Revision der Produkthaftungsrichtlinie v. 28.9.2022, COM (2022) 495 final, Art. 6, S. 31 f. und passim.

85 *G. Wagner*, Roboter als Haftungssubjekte? Konturen eines Haftungsrechts für autonome Systeme, in: *F. Faust/ H. Schäfer* (Hrsg.), Zivilrechtliche und rechtsökonomische Probleme des Internet und der künstlichen Intelligenz, 2019, S. 1 (18 ff.); *ders.*, Verantwortlichkeit (Fn. 12), 734 f.

86 So das Verständnis des EuGH, vgl. EuZW 2002, 574; ausf. zum Konzept der Vollharmonisierung *T. Riehm*, 25 Jahre Produkthaftungsrichtlinie, EuZW 2010, 567; *R. Schaub*, Europäische Produkthaftung: Wie weit reicht die Harmonisierung heute?, ZEuP 2011, 41.

87 Ökonomisch betrachtet sind diejenigen Schäden zu vermeiden, deren Kosten (Ersatz) höher sind als die Kosten, die für ihre Vermeidung hätten aufgewandt werden müssen, grdl. *S. Shavell*, *Foundation of Economic Analysis of Law*, Cambridge, Mass.: Harvard University Press 2004, S. 178 f.

Hauptzweck des Haftungsrechts und in einer Gefährdungshaftung das beste Instrument zur Steuerung des Aktivitätsniveaus des Herstellers, mithin zur Entwicklung von Sicherheitstechnologien.⁸⁸

Doch dürfte das einzig probate Mittel zur Verhaltenssteuerung des Herstellers im *Fehlererfordernis* liegen.⁸⁹ Anders als im System der Verhaltenspflichten, die mehr oder weniger klar am Stand der Technik orientiert sind, fehlt bei der Gefährdungshaftung ein wirklich handhabbarer Orientierungspunkt für unternehmerische Vorsorgemaßnahmen. Es fällt dem Unternehmer schwer, die Kosten von Sicherheitsmaßnahmen mit denen möglicher Schäden abzuwägen, wenn die Schäden gerade KI-spezifisch unvorhersehbar sind und auch für unvorhersehbare Schäden gehaftet wird.⁹⁰ Zudem kann sich eine strikte Haftung für neue Technologien als echter Innovationshemmer erweisen, wenn selbst für rechtmäßiges Verhalten gehaftet wird.⁹¹ Schließlich bestehen auch dogmatische Bedenken gegenüber einer Gefährdungshaftung. Die Eigenheiten des digitalen *Autonomierisikos* passen nicht auf die Gefährdungshaftung. Bei KI kommt es „gerade nicht auf die Sachgefahr eines falsch funktionierenden Computers, also das Kausalrisiko an, sondern auf das Entscheidungsrisiko, auf die andersgeartete Gefahr, dass sich dessen autonome Entscheidungen als Fehlentscheidungen herausstellen. Zurechnungsgrund ist nicht der Einsatz eines Objekts erhöhter Gefahr, sondern das rechtswidrige Verhalten des Algorithmus, den der Prinzipal rechtmäßig zu eigenen Nutzen eingesetzt hat.“⁹² Autonomiegefahr kann nicht mit Betriebsgefahr gleichgesetzt werden.

88 So etwa *Wagner*, *Roboter* (Fn. 85), 7 ff.; *ders.*, *Verantwortlichkeit* (Fn. 12), 722.

89 *Denga*, *Deliktische Haftung* (Fn. 45), 76; *Brand*, *Haftung* (Fn. 7), 948; *C. Picker*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum 2020* (Fn. 22), 760; *Thöne*, *Systeme* (Fn. 16), S. 167 ff., 180.

90 *Denga*, *Deliktische Haftung* (Fn. 45), 76; ähnlich *Mühlböck/Taupitz*, *Haftung* (Fn. 7), 212.

91 Kritisch auf dem *Karlsruher Forum 2020* u.a. *D. Looschelders*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum 2020* (Fn. 22), 757: ungewollte Haftungsausdehnung; *T. Lobinger*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum 2020* (Fn. 22), 760; *F. Maultsch*, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum 2020* (Fn. 22), 758 f.: Innovationsmalus.

92 *Teubner*, *Rechtssubjekte* (Fn. 12), 192.

2. Verschuldensunabhängige Unrechtshaftung

Sachgerecht ist weiterhin die verschuldensunabhängige Unrechtshaftung, allerdings mit einer *Umkehr der Beweislast* bezüglich des Produktfehlers.⁹³ Bei der Haftung nach ProdHaftG muss der europäische Gesetzgeber tätig werden, weil Art. 4 der Richtlinie dem Geschädigten die Beweislast für den Produktfehler auferlegt.⁹⁴ Bei § 823 Abs. 1 BGB lässt sich das durch die Rechtsprechung realisieren. Im Falle legislativer Untätigkeit sollte sie für Digitalprodukte die Beweislast nicht nur hinsichtlich des Verschuldens des Herstellers, sondern auch für den Produktfehler umkehren.⁹⁵

Wegen des angesprochenen Vernetzungsrisikos sind zudem *Beweiserleichterungen* bezüglich der *Kausalität* zu gewähren.⁹⁶ Die konstante Vernetzung mit anderen, von Dritten kontrollierten KI-Systemen oder Datenquellen führt zu Problemen, eine für den konkreten Schaden verantwortliche Person

93 So auch *Expert Group on Liability and New Technologies*, Liability (Fn. 58), S. 55. Die Bundesregierung äußert in ihrer Stellungnahme zum Weißbuch v. 29.6.2020, S. 26, hingegen Zweifel an der Erforderlichkeit einer Beweislastumkehr. Der Richtlinienentwurf zur Revision der Produkthaftungsrichtlinie v. 28.9.2022 sieht keine Beweislastumkehr, doch Beweiserleichterungen und Modifikationen des Beweismaßes vor, COM (2022) 495 final, Art. 9, S. 33 f.

94 Da es sich nach Ansicht des EuGH bei der Produkthaftungsrichtlinie um einen Akt der Vollharmonisierung handelt, kann von ihren Vorgaben nicht abgewichen werden, vgl. EuZW 2002, 574; Katzenmeier/Voigt (Fn. 55), Einl. Rn. 5; ausf. zum Konzept der Vollharmonisierung bei der Produkthaftung Riehm, Produkthaftungsrichtlinie (Fn. 86), 567 ff.; Schaub, Produkthaftung (Rn. 86), 41 ff.

95 So auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss Nr. 1; Wagner, Verantwortlichkeit (Fn. 12), 735, gefordert als Minderform der von ihm bejahten Gefährdungshaftung; Zech, Entscheidungen (Fn. 28), S. A 59, A 73, A 86; auch mit Blick auf nationale Haftungsvorschriften EU-Kommission, Bericht zum Weißbuch v. 19.2.2020 – COM (2020) 64 final, S. 17; krit. aber Bundesregierung, Stellungnahme zum Weißbuch v. 29.6.2020, S. 27.

96 So auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss Nr. 2. Bisweilen wird die Kausalitätsvermutung des § 84 Abs. 2 AMG, die Schwierigkeiten des Ursächlichkeitsnachweises bei Summations- und Distanzschäden begegnen soll, als beispielhafte Regelung angeführt. Die Rechtsnatur ist allerdings umstritten und von der Rechtsprechung wird die Norm restriktiv ausgelegt, vgl. BGH MedR 2013, 729 m. krit. Anm. D. Hart, Über den Umgang des VI. Senats des BGH mit der Reform der Arzneimittelhaftung, MedR 2013, 705; eingehend ders., in: H. Rieger/F. Dahm/C. Katzenmeier/M. Stellpflug/O. Ziegler, HK-AKM, 63. Lfg., Heidelberg 2016, Arzneimittelhaftung, Nr. 243, Rn. 66 ff.; C. Katzenmeier/T. Voigt, Das Beweisrecht der Produkthaftung unter europäischem Einfluss, in: C. Althammer/C. Schärfl (Hrsg.), Dogmatik als Fundament für Forschung und Lehre – Festschrift für Herbert Roth zum 70. Geburtstag, Tübingen 2021, S. 947 (967 ff., 970 f.); Voigt, Produkthaftung (Fn. 17), § 3 B. I. 2. b) u. II. 3. zu § 84 Abs. 2 AMG, § 4 B. II. zu dessen Europarechtskonformität und § 5 B. IV. zu jüngeren

zu identifizieren. Sie wird das Haftungsrecht vor bislang nicht bekannte Schwierigkeiten stellen.⁹⁷ In ersten Stellungnahmen wird eine „systemische Haftung“ aller an der Vernetzung Beteiligten gefordert,⁹⁸ die Verantwortung einer „anonymen Matrix“ von sozialen und digitalen Handlungen (Risikopools) in Betracht gezogen,⁹⁹ die Definition produktübergreifender Systeme und Haftung des „Systemadministrators“ für Systemversagen erwogen¹⁰⁰ oder eine Ersetzung oder Ergänzung der Individualhaftung durch eine Versicherungslösung für notwendig erachtet.¹⁰¹ Gewisse Vergünstigungen kann der nach geltendem Recht beweisbelastete Geschädigte erfahren, wenn Herstellern die Pflicht auferlegt wird, in das autonome System ein Gerät zu installieren, das datenschutzkonform gewisse Parameter misst und aufzeichnet, um so die Entscheidung des autonomen Systems im Nachhinein nachvollziehen zu können und die Beweissicherung zu gewährleisten.¹⁰²

Verallgemeinerungstendenzen – Die *EU-Kommission* erwägt in ihrem Bericht zum Weißbuch v. 19.2.2020 – COM (2020) 64 final, S. 20 eine Beweislastumkehr hinsichtlich der Kausalität; s. auch *Wagner*, Verantwortlichkeit (Fn. 12), 730; s. weiter den Richtlinienentwurf zur Revision der Produkthaftungsrichtlinie v. 28.9.2022, COM (2022) 495 final, Art. 9 Nr. 3 u. 4, S. 34; zu einer Modifikation des allgemeinen Haftungsrechts bzgl. KI s. Richtlinienentwurf v. 28.9.2022, COM (2022) 496 final, Art. 4, S. 31 ff. und passim.

97 *Teubner*, Rechtssubjekte (Fn. 12), 158, 201; *Wagner*, Verantwortlichkeit (Fn. 12), 720 f.

98 *Spiecker gen. Döhmman*, Zur Zukunft systemischer Digitalisierung, CR 2016, 698 (703); krit. ggü. Beweismaßreduzierungen und einer probabilistischen Proportionalhaftung *C. Katzenmeier*, Beweismaßreduzierung und probabilistische Proportionalhaftung, ZZP 117 (2004), 187; *G. Spindler*, Kausalität im Zivil- und Wirtschaftsrecht, AcP 208 (2008), 283 (305 ff.).

99 *Teubner*, Rechtssubjekte (Fn. 12), 158, 202 f.

100 *Wagner*, Verantwortlichkeit (Fn. 12), 733, 740.

101 S. u. C. III.; auch bereits *C. Katzenmeier*, Entwicklungen des Produkthaftungsrechts, JuS 2003, 943 (950 f.).

102 *L. Lutz*, Autonome Fahrzeuge als rechtliche Herausforderung, NJW 2015, 119 (120); *S. Horner/M. Kaulartz*, Haftung 4.0, CR 2016, 7 (10); *Schaub*, Interaktion (Fn. 16), 344 f.; *M. Lohmann*, Roboter als Wundertüten, AJP 2017, 152 (158); s. auch *EU-Kommission*, Bericht zum Weißbuch v. 19.2.2020 – COM (2020) 64 final, S. II. – § 63a StVG, eingeführt durch Gesetz v. 16.6.2017, BGBl. I 1648, regelt die Datenverarbeitung bei Kfz mit hoch- oder vollautomatisierter Fahrfunktion, Abs. 3 die Übermittlung der Daten bei der Geltendmachung von Schadensersatzansprüchen.

IV. Haftungsadressat

Zweite zentrale Frage ist die Verteilung der Haftung zwischen dem Hersteller und dem Anwender eines autonomen Systems.¹⁰³

1. Hersteller

Tendenziell wird eher eine Haftung des Herstellers in Betracht gezogen und dafür der Gesichtspunkt der *Risikokontrolle* angeführt. Die Devise lautet: Haftung folgt Kontrolle, ohne Kontrolle keine Haftung.¹⁰⁴ Der Hersteller des autonomen Systems ist diejenige Partei, die das Verhalten des Produkts in dem Umfang determiniert, wie dies überhaupt möglich ist.¹⁰⁵

Dabei reicht sein Einfluss bei digitalen Systemen anders als bei analogen Produkten regelmäßig über den Zeitpunkt des Inverkehrbringens hinaus, erstreckt sich auf die volle Lebensdauer des Produkts. So verändert der Hersteller durch *Software-Updates* die Steuerung des digitalen Geräts.¹⁰⁶ Deswegen ist die Fixierung der Herstellerpflichten auf den Zeitpunkt des Inverkehrbringens in § 1 Abs. 2 Nr. 2 ProdHaftG bei digitalen Produkten nicht angemessen,¹⁰⁷ Art. 6 Abs. 1 lit. c) ProdHaft-RI ist entsprechend zu ändern.¹⁰⁸

103 D. Looschelders, zit. nach *Biller-Bomhardt/Kunz*, *Karlsruher Forum* 2020 (Fn. 22), 755; s. auch *Schaub*, *Interaktion* (Fn. 16), 348; *Thöne*, *Systeme* (Fn. 16), S. 161 ff.

104 *Wagner*, *Verantwortlichkeit* (Fn. 12), 724 f.

105 *Horner/Kaulartz*, *Haftung* (Fn. 102), 9; *Wagner*, *Verantwortlichkeit* (Fn. 12), 734.

106 *Wagner*, *Verantwortlichkeit* (Fn. 12), 734. Unterlässt der Nutzer die Installation nötiger Updates, kommt eine Haftungsminde rung in Betracht. Modifiziert der Anwender ein offenes System (z.B. durch Aufspielen eines eigenen Programms), übernimmt er Kontrolle und damit Verantwortung.

107 *Wagner*, *Verantwortlichkeit* (Fn. 12), 728.

108 Vgl. den Richtlinienentwurf zur Revision der Produkthaftungsrichtlinie v. 28.9.2022, COM (2022) 495 final, Art. 6 Nr. 1 lit. c)–f), S. 31 f. und passim. Nach Umsetzung der Richtlinie über digitale Inhalte und Dienstleistungen (DIDRL – (EU) 2019/770) trifft ab dem 1.1.2022 (Ges. v. 25.6.2021, BGBl. I 2126) den Verkäufer von Waren mit digitalen Elementen gemäß § 327f Abs. 1 S. 1 BGB eine Aktualisierungspflicht für die digitalen Elemente. Erfolgt der Verkauf nicht durch den Hersteller, so ist der Verkäufer auf die Mitwirkung des Herstellers angewiesen und muss dies frühzeitig vertraglich sicherstellen. Auf Behandlungsverträge nach § 630a BGB finden die Vorschriften über digitale Inhalte nach § 327 Abs. 6 Nr. 3 BGB keine Anwendung. – Für eine allg. deliktsrechtliche Updatepflicht B. *Raue*, *Haftung für unsichere Software*, *NJW* 2017, 1841 (1844); *Wagner* (Fn. 30), § 823 Rn. 1008 ff.; *ders.*, *Produkthaftung* (Fn. 5), 755 ff.; *ders.*, *Verantwortlichkeit* (Fn. 12), 728; *Zech*, *Entscheidungen* (Fn. 28), S. A 50; dagegen *Spindler* (Fn. 25), § 823 Rn. 761; *ders.*, *Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz*, *CR* 2015, 766 (770); *P. Schrader/J. Engstler*, *Anspruch auf Bereit-*

Kommt es für die Beurteilung der Fehlerhaftigkeit des digitalen Produkts auf den Zeitpunkt des jeweiligen (auch freiwilligen) Updates an, so muss zur Vermeidung einer zeitlich unbegrenzten Herstellerhaftung für den Beginn der Verjährung weiterhin der Zeitpunkt des Inverkehrbringens maßgebend sein.¹⁰⁹

Es bleibt die Frage, ob der Hersteller notwendigerweise Zentralfigur des Haftungsgeschehens ist.¹¹⁰ Der These, dass die Haftung der Kontrolle folgen müsse, liegt die (Über-) Betonung der Präventivfunktion zugrunde.

2. Anwender

Für eine Haftung des Anwenders spricht, dass dieser den konkreten Nutzen aus dem Einsatz des autonomen Systems zieht. Er entscheidet über den zeitlichen und räumlichen Einsatz des Systems sowie Art und Weise der Benutzung.¹¹¹

Hinzu kommt, dass bei hochkomplexen und vernetzten Systemen der Anwender leichter auszumachen ist als der Hersteller der schadenstiftenden Komponente. Bei Schäden im Zuge einer medizinischen Behandlung ist in aller Regel der Arzt oder Krankenträger „Anlaufstelle“ des anspruchstellenden Patienten. Unterwirft man ihn einer strengen Haftung, etwa indem man doch den Einsatz des autonomen Systems per se für pflichtwidrig erachtet,¹¹² verlagert man manche Zweifelsfrage in das Regressverhältnis gegenüber dem Hersteller. Die Haftung könnte ähnlich der bei Verkehrsunfällen ausgestaltet werden.¹¹³ Dort wird es dem Geschädigten nicht zugemutet,

stellung von Software-Updates?, MMR 2018, 356 (360); H. Wiesemann/C. Mattheis/S. Wende, Software-Updates bei vernetzten Geräten, MMR 2020, 139 (140); Voigt, Produkthaftung (Fn. 17), § 5 B. III. 2.

109 G. Spindler, Haftung für autonome Systeme – ein Update, in: S. Beck/C. Kusche/B. Valerius (Hrsg.), Digitalisierung, Automatisierung, KI und Recht, Baden-Baden 2020, S. 255 (263).

110 Zumal es „den“ Hersteller unter Umständen nicht mehr gibt; in Betracht kommen dann Hersteller verschiedener Teilkomponenten des KI-Systems, Entwickler des Codes, Lieferanten der Daten für das Training der KI sowie Betreiber von Netzwerken, über die die KI eingespielt und aktualisiert wird, vgl. Spindler, in: C. Katzenmeier (Hrsg.), FS Hart (Fn. 12), S. 581 (594).

111 T. Riehm, zit. nach Biller-Bomhardt/Kunz, Karlsruher Forum 2020 (Fn. 22), 760; Zech, Entscheidungen (Fn. 28), S. A 88, zur Verantwortung professioneller Betreiber S. A 100.

112 Dagegen sub. B. I. 1. a) und 2. a).

113 Dafür 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschlüsse Nr. 13 b), 14 c), 15 b).

sich wegen seiner Schadensersatzansprüche an den zuständigen Kfz-Hersteller zu halten.¹¹⁴ Der Halter haftet streng gemäß § 7 Abs. 1 StVG (auch für Schäden durch vollautomatisierte Kfz), die haftungsrechtliche Verantwortlichkeit des Herstellers wird regelmäßig erst im Wege des Regresses der Haftpflichtversicherung des Halters erreicht.¹¹⁵

Fazit: Gestützt auf Ausgleichs- und Präventionsfunktion müssen die Aspekte Risikosteuerung, Vorteilsziehung und Zurechenbarkeit gewichtet werden, um das Haftungsrisiko zwischen dem Hersteller des autonomen Systems und dem Anwender angemessen zu verteilen.¹¹⁶

3. Anerkennung einer e-Person?

Wegen der zunehmenden Intelligenz und Autonomie von KI-Systemen und der mit deren Einsatz verbundenen rechtsdogmatischen Schwierigkeiten sind Rufe nach einer Gesetzesänderung zur Anerkennung der *Rechtsfähigkeit* bestimmter autonomer Systeme zu vernehmen.¹¹⁷ Das mag insbesondere bei anthropomorphen Robotern wie dem Pflegeroboter *Pepper* intuitiv naheliegen.¹¹⁸ Vorteil der *elektronischen Person (e-Person)*: Gerade wenn sich die Schädigung beim Einsatz eines autonomen Systems nicht auf einen bestimmten Akteur zurückführen lässt, könnte in der Promotion des autonomen Systems zum Haftungssubjekt die Lösung aller Zurechnungsprobleme liegen.¹¹⁹

114 Hinweis von *Wagner*, Verantwortlichkeit (Fn. 12), 731 f.

115 *Wagner*, Verantwortlichkeit (Fn. 12), 738: Die Inanspruchnahme des „richtigen“ Haftungssubjekts im Regressweg gemäß § 86 VVG, also „übers Eck“, ist in administrativer Hinsicht freilich die aufwendigere und kostenintensivere Lösung.

116 *Schaub*, Interaktion (Fn. 16), 346; *Zech*, Entscheidungen (Fn. 28), S. A 88 ff.

117 Zuerst wohl *A. Matthias*, Automaten als Träger von Rechten, Berlin 2008 / 2. Aufl., Berlin 2010, S. 83 ff. u. 111 ff.; *S. Beck*, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009, 225 (229 f.); *E. Hilgendorf*, Können Roboter schuldhaft handeln?, in: *S. Beck* (Hrsg.), *Jenseits von Mensch und Maschine*, Baden-Baden 2012, S. 119 (125 ff.); *S. Beck*, Über Sinn und Unsinn von Statusfragen, in: *E. Hilgendorf / J. Günther* (Hrsg.), *Robotik und Gesetzgebung*, Baden-Baden 2013, S. 239 (255 ff.); *J. Günther*, *Roboter und rechtliche Verantwortung*, München 2016, 251 ff. – dazu monographisch *Linardatos*, Aktanten (Fn. 17).

118 *T. Riehm*, Nein zur ePerson!, RD 2020, 42; *Wagner*, Verantwortlichkeit (Fn. 12), 738; *Thöne*, Systeme (Fn. 16), S. 182.

119 *Teubner*, Rechtssubjekte (Fn. 12), 157; *Linke*, Elektronische Person (Fn. 39), 200: auf den ersten Blick eine einfache „one-fits-all“-Lösung.

Die vormalige Aufforderung des Europäischen Parlaments an die Kommission, bei der Suche nach künftigen Rechtsinstrumenten auch in ihre Erwägungen mit einzubeziehen, ob für die „ausgeklügelten autonomen Roboter“ langfristig ein spezieller rechtlicher Status zu schaffen ist,¹²⁰ hat indes breiten Widerspruch ausgelöst.¹²¹ Teilweise werden fundamentale epistemologische und ontologische Bedenken geäußert und argumentiert, dass Rechtssubjektivität allein Menschen vorbehalten sei.¹²² Rechtsfähigkeit ist indes nicht auf natürliche Personen beschränkt, wie die juristische Person zeigt.¹²³ Und für das Haftungsrecht ist auch keine volle Rechtssubjektivität erforderlich, vielmehr eine funktionsbezogene *Teilrechtsfähigkeit* ausreichend, um Adressat von Schadensersatzansprüchen sein zu können.¹²⁴ Insofern ist zu fragen, welche Folgen die Anerkennung der e-Person für die haftungsrechtlichen Zwecke Schadensausgleich und Prävention hätte.¹²⁵

Da e-Personen über keine Vermögenswerte verfügen, aus denen Schadensersatzansprüche befriedigt werden können, müssten sie mit einer Haftungs-*masse* ausgestattet werden. Diese wäre von denjenigen aufzubringen, die sonst persönlich für die vom Roboter verursachten Schäden haften müssten,

120 *EU-Parlament*, Entschließung v. 16.2.2017 – 2015/2013 (INL), Rn. 59 sub lit. f.

121 Vgl. nur etwa *Expert Group on Liability and New Technologies*, Liability (Fn. 58), S. 37 ff.; *Datenethikkommission*, Gutachten (Fn. 16), S. 219.

122 *H. Eidenmüller*, The Rise of Robots and the Law of Humans, ZEuP 2017, 765 (775 f.); *C. Müller-Hengstenberg/S. Kirn*, Intelligente (Software-)Agenten, MMR 2014, 307 (307 f.); s. auch *Brand*, Haftung (Fn. 7), 948: Risiko, ein neues „Untermenschentum“ heraufzubeschwören; *Mühlböck/Taupitz*, Haftung (Fn. 7), 214.

123 Die Rechtsfähigkeit ist jeweils einfachgesetzlich anerkannt, vgl. § 21 BGB, § 1 Abs. 1 S. 1 AktG, § 13 Abs. 1 GmbHG, § 17 Abs. 1 GenG, Art. 1 Abs. 3 SE-VO. Juristische Personen handeln allerdings nicht selbständig, sondern werden durch Menschen vertreten.

124 In diesem Sinne *M. Gruber*, Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs, in: Beck (Hrsg.), Mensch und Maschine (Fn. 117), S. 133 (155 ff.); *G. Teubner*, Elektronische Agenten und große Menschenaffen, ZfRS 2006, 5 (29 f.); *ders.*, Rechtssubjekte (Fn. 12), 163 ff.; für eine Teilrechtsfähigkeit im vertraglichen, nicht aber im deliktischen Bereich auch *Schirmer*, Roboter (Fn. 35), 663 ff.; *ders.*, Von Mäusen, Menschen und Maschinen, JZ 2019, 711 (716); abl. etwa *S. Klingbeil*, Der Begriff der Rechtsperson, AcP 217 (2017), 848 (859 f.); *ders.*, Schuldnerhaftung für Roboterversagen, JZ 2019, 718 (722); *Riehm*, ePerson (Fn. 118), 46 ff.; *Voigt*, Produkthaftung (Fn. 17), § 5 C. I. 2. b).

125 *Wagner*, Verantwortlichkeit (Fn. 12), 738; *ders.*, Haftungssubjekte (Fn. 85), S. 29 ff.; s. auch *J. Kersten*, Menschen und Maschinen Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen, JZ 2015, 1 (6 f.); *Schirmer*, Roboter (Fn. 35), 663; *Teubner*, Rechtssubjekte (Fn. 12), 163.

also insbesondere Hersteller und Anwender.¹²⁶ Im wirtschaftlichen Ergebnis wird mit Anerkennung der e-Person die Haftung der sonst verantwortlichen Akteure auf die Versicherungssumme des Roboters reduziert.¹²⁷ Es stünde zu befürchten, dass die Nutzer autonome Systeme als Haftungsschild missbrauchen.¹²⁸ Und für ökonomische Anreize ist ein digitales System nicht empfänglich. Aus haftungsrechtlicher Sicht ist die Idee der e-Person deshalb nicht nur sinnlos, sondern kontraproduktiv.¹²⁹

V. Versicherungslösung

Deutlich wird, dass gewisse Schwierigkeiten bestehen, Schäden infolge des Einsatzes autonomer Systeme in der Medizin mit dem geltenden Haftungsrecht angemessen zu verarbeiten. Dies kann dazu veranlassen, insoweit über einen Systemwechsel nachzudenken, und zwar über eine Ablösung der geltenden Individualhaftung durch ein *kollektives Entschädigungssystem* nach dem Vorbild der gesetzlichen Unfallversicherung, §§ 104 ff. SGB VII¹³⁰ oder eine Ergänzung der Individualhaftung durch Einrichtung eines *Entschädigungsfonds*.¹³¹ Dem Patienten würde ein genuiner Entschädigungsanspruch gewährt, der unabhängig ist von der Haftung. Ein Nachweis der Verursa-

126 Wagner, Verantwortlichkeit (Fn. 12), 739; ders., Haftungssubjekte (Fn. 85), S. 32; Zech, Entscheidungen (Fn. 28), S. A 97; Riehm, ePerson (Fn. 118), 46.

127 Wagner, Verantwortlichkeit (Fn. 12), 739. Dann aber läge es näher, an eine neue Gesellschaftsform („e-GmbH“) zu denken. Effizienter als die e-Person wäre eine Haftpflichtversicherung mit Mindestdeckungssumme.

128 Denga, Deliktische Haftung (Fn. 45), 77; G. Spindler, Privatrechtsdogmatik und Herausforderungen der IT-Revolution, in: H. Grigoleit/J. Petersen (Hrsg.), Privatrechtsdogmatik im 21. Jahrhundert, Festschrift für Claus-Wilhelm Canaris zum 80. Geburtstag, Berlin Boston 2017, S. 709 (714); Linke, Elektronische Person (Fn. 39), 202.

129 Wagner, Verantwortlichkeit (Fn. 12), 739; ders., in: Haftungssubjekte (Fn. 85), S. 32 f.; Spindler, Digitale Wirtschaft – analoges Recht (Fn. 16), 816; Thöne, Systeme (Fn. 16), S. 182 f.; Voigt, Produkthaftung (Fn. 17), § 5 C. I. 2. b); abl. auch 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschluss 16.

130 Bei fortschreitender Vernetzung befürwortend Zech, Entscheidungen (Fn. 28), S. A 107 ff.; ders., Intelligenz (Fn. 13), 208 f., 216 ff.; für bestimmte Bereiche wie die Medizintechnik erwägend Janal, deliktische Haftung (Fn. 81), S. 139 (157). Zu Überlegungen einer allg. Ersetzung der Arzthaftung durch Versicherungsschutz s. C. Katzenmeier, Arzthaftung, Tübingen 2002, S. 214 ff.; ders., „Heilbehandlungsrisikoversicherung“- Ersetzung der Arzthaftung durch Versicherungsschutz?, VersR 2007, 137.

131 Für eine allg. Ergänzung der Arzthaftung D. Hart/R. Francke, Der Modellversuch eines Patientenentschädigungs- und -härtefallfonds, 2013; dazu C. Katzenmeier, Patientenentschädigungsfonds – rechtspolitische Forderungen und rechtsdogmatische

chung durch einen bestimmten Akteur wäre mit einer solchen Versicherungslösung obsolet und die Schadensabwicklung würde stark vereinfacht.¹³²

Wo das Haftpflichtrecht zur Bewältigung eines durch die soziale und technische Entwicklung gestiegenen Gefahrenpotentials nicht ausreicht, sind Versicherungskonzepte in Betracht zu ziehen als eine Möglichkeit, dem gewachsenen Sekuritätsbedürfnis Rechnung zu tragen und den für erforderlich gehaltenen Schutz der Verkehrsteilnehmer zu gewähren. Es ist einstweilen nicht zu sehen, wie die Fälle multipler Kausalität infolge Vernetzung digitaler Produkte mit den Mitteln des geltenden Rechts (§§ 830, 840 BGB) bewältigt werden können. Eine Versicherungslösung zur Bewältigung von Schäden durch autonome Systeme wirft indes ihrerseits schwierige Fragen auf. So bereitet schon eine überzeugende Bestimmung des modalen Schutzbereichs Schwierigkeiten, sodann führen anonyme Versicherungslösungen und Entschädigungspools unweigerlich zu einer Verdünnung von Verantwortlichkeiten, es kommt zu Präventionseinbußen, auch sind die Finanzierungsfragen gänzlich ungeklärt.¹³³

D. Fazit

Autonomie- und Vernetzungsrisiken von KI-Systemen fordern das Recht heraus. Dieses hat die schwierige Aufgabe, Innovationsoffenheit und Innovationsverantwortung zu justieren.¹³⁴ Die medizinische Anwendung von KI stellt quasi die Probe aufs Exempel dar, denn hier zeigt sich besonders deutlich,

Erwägungen, VersR 2014, 405; G. Wagner, Bedarf es einen Härtefallfonds für Behandlungsschäden?, MedR 2021, 101.

132 Katzenmeier, Arzthaftung (Fn. 130), S. 218 (244); ders., Heilbehandlungsrisikoversicherung (Fn. 130), 137, 139 f.; Mühlböck/Taupitz, Haftung (Fn. 7), 213.

133 Vgl. Katzenmeier (Fn. 23), Kap. X Rn. 133 (Haftungersetzung) u. 157 ff. (Haftungsergänzung); spez. im vorliegenden Kontext krit. Brand, Haftung (Fn. 7), 947; Wagner, Verantwortlichkeit (Fn. 12), 740 f.; Mühlböck/Taupitz, Haftung (Fn. 7), 213; auf dem Karlsruher Forum 2020 wurde eine Unfallversicherung nicht als geeignet erachtet zur Erfassung von Risiken digitaler Techniken, s. Tagungsbericht Biller-Bomhardt/Kunz, Karlsruher Forum 2020 (Fn. 22), 755.

134 W. Hoffmann-Riem, Innovation und Recht – Recht und Innovation. Recht im Ensemble seiner Kontexte, Tübingen 2016, S. 28 ff.; ders., Verhaltenssteuerung durch Algorithmen – Eine Herausforderung für das Recht, AöR 142 (2017), 1, 5; ders., Innovationsoffenheit und Innovationsverantwortung durch Recht: Aufgaben rechtswissenschaftlicher Innovationsforschung, AöR 131 (2006), 255; Katzenmeier, Big Data (Fn. 1), 259, 262 u. 271.

welch enormes Potential KI für die Diagnose und Therapie wie Prävention von Krankheiten hat, aber auch welche Gefahren und Bedenken bestehen.¹³⁵

Die Haftung des Herstellers autonomer Systeme ist nicht als reine Gefährdungshaftung auszugestalten, vielmehr als verschuldensunabhängige Fehlerhaftung. Die Beweislast bezüglich des Fehlers ist umzukehren, bezüglich der Kausalität sind Beweiserleichterungen zu gewähren. Der Anwender haftet für eigene Fehler beim Einsatz von KI, bei Versagen des maschinellen Verrichtungsgehilfen wird das Auswahl- und Überwachungsver schulden vermutet. Die Exkulpationsmöglichkeit gemäß § 831 Abs. 1 S. 2 BGB ist zu streichen – sie ist selbst beim menschlichen Verrichtungsgehilfen, der seinerseits haftet, verfehlt.¹³⁶

Angezeigt scheint überdies die Anordnung einer *Pflicht-Haftpflichtversicherung* des Herstellers wie des Anwenders digitaler Produkte, jedenfalls bei erheblichem Risikopotential.¹³⁷ Wird dem Geschädigten zudem nach dem Vorbild der Kfz-Haftpflichtversicherung ein *Direktanspruch* gegen den Versicherer eingeräumt, so erhält er einen solventen Schuldner und ist besser geschützt gegen Einwendungen des Versicherers aus dem Versicherungsverhältnis, z.B. wegen Prämienzahlungsverzugs oder Obliegenheitsverletzungen des Versicherungsnehmers.¹³⁸ Auch dann bleibt die Haftpflichtversicherung freilich der Haftung nachgeschaltet.¹³⁹

Nicht zu verkennen ist, dass Patientenschäden infolge des Einsatzes autonomer Systeme auch ohne eine spezielle Versicherungslösung bereits *de lege lata* weitgehend *kollektiviert* sind, nämlich durch die Systeme sozialer Siche-

135 Katzenmeier, Big Data (Fn. 1), 259, 268 f.; A. Lohmann/A. Schömig, „Digitale Transformation“ im Krankenhaus. Gesellschaftliche und rechtliche Herausforderungen durch das Nebeneinander von Ärzten und Künstlicher Intelligenz, in: S. Beck/C. Kusche/B. Valerius (Hrsg.), Digitalisierung, Automatisierung, KI und Recht, Baden-Baden 2020, S. 345 ff.; Mühlböck/Taupitz, Haftung (Fn. 7), 182, 216.

136 Nachweise zu Reformbestrebungen bei Bernau (Fn. 43), § 831 Rn. 187 ff.

137 Dafür EU-Parlament, Entschließung v. 16.2.2017 – 2015/2103 (INL), Rn. 57; Spindler, Roboter (Fn. 108), 775; G. Wagner, Robot Liability, in: Lohsse et. al. (Hrsg.), Liability (Fn. 81), S. 27 (58 f.); G. Borges, New Liability Concepts: the Potential of Insurance and Compensation Funds, in: Lohsse et. al. (Hrsg.), Liability (Fn. 81), S. 145, 153 ff.; Thöne, Systeme (Fn. 16), S. 261 ff.; diff. 73. DJT, Bonn 2022, Abt. Zivilrecht, Beschlüsse 17 u. 18.

138 § 117 VVG mit Regelung der Nachhaftung in Abs. 2 und des Umfangs der Leistungspflicht in Abs. 3.

139 S. aber zu Rückwirkungen der Pflichthaftpflichtversicherung auf die Haftung C. Armbrüster, Auswirkungen von Versicherungsschutz auf die Haftung, NJW 2009, 187; monographisch M. Makowsky, Der Einfluss von Versicherungsschutz auf die außervertragliche Haftung, Karlsruhe 2013.

140 Besteht Behandlungsbedarf, dann erbringen diese Leistungen. Schadensersatzansprüche des Patienten gehen insoweit gemäß § 116 Abs. 1 SGB X kraft Gesetzes auf sie über (bei Privatversicherung gemäß § 86 VVG, bei Berufsbeamten gemäß § 76 BBG). Für den im Regressweg in Anspruch genommenen Schädiger (insbesondere Hersteller, unter Umständen Anwender des autonomen Systems) tritt dessen Haftpflichtversicherer ein. Die Schadensregulierung erfolgt zwischen den Kollektiven: Krankenversicherer versus Haftpflichtversicherer, hier zumeist nicht im Einzelfall, vielmehr durch Regressverzichts- und Schadensteilungsabkommen. Eine Verschärfung der Haftungsregeln verschiebt die ausgehandelte Quote zugunsten der Systeme sozialer Sicherung.¹⁴¹

Da der Einsatz von KI einen Zugewinn an Sicherheit und Qualität der Krankenversorgung verspricht, wird erwartet, dass es insgesamt zu weniger Schadensfällen kommt, mithin weniger Haftungsprozesse ausgetragen und die Versicherungssysteme entlastet werden. Nicht weniger problematisch aber wird – auch ungeachtet aller Dystopien eines posthumanen Zeitalters¹⁴² – der mit wachsender Autonomie der Systeme zunehmende Kontrollverlust des Menschen über die Maschinen.

140 Vgl. Katzenmeier, *Arzthaftung* (Fn. 130), S. 204.

141 C. Katzenmeier, *Überlagerung des Schadensrechts durch das Versicherungsrecht*, *VersR* 2002, 1449 (1454); *ders.* (Fn. 41), *Vorbem. zu §§ 823 ff. Rn. 47 ff.*; *spez. zur Arzthaftung ders.* (Fn. 23), *Kap. X Rn. 128.*

142 S. nur etwa N. Bostrom, *Superintelligenz. Szenarien einer kommenden Revolution*, Berlin 2014.

Google, Facebook & Co: Die Macht von Daten und Algorithmen im Fokus des Kartellrechts*

Torsten Körber

A. Ausgangspunkt

Wer an die „Macht der Algorithmen“ – so der Titel der Ringvorlesung – denkt, der kommt nahezu zwangsläufig auf die großen Internetkonzerne Google (bzw. Alphabet), Amazon, Facebook (bzw. Meta), Apple und Microsoft – kurz: „GAFAM“. Machtkonzentrationen sind in der Wirtschaft nichts Neues. Schon in den 1880er Jahren wurde in den USA die Sorge laut, dass Konzentrationen auf dem Gebiet der Eisenbahn und der Öl-Industrie, z.B. in Gestalt des „Standard Oil Trusts“ zu wirtschaftlich und politisch unkontrollierbaren Monopolen führen könnten. Damals wurde als Antwort darauf im Jahre 1890 in den USA mit dem Sherman Act das erste moderne Kartellgesetz der Welt eingeführt, das bis heute besteht und Kartelle ebenso verbietet wie eine Monopolisierung mit wettbewerbswidrigen Mitteln. Regelungen gegen Kartelle und Machtmissbräuche finden sich heute in praktisch jeder Kartellrechtsordnung und sie haben in den vergangenen 130 Jahren gute Dienste für den Wettbewerb geleistet. Die Mineralölmärkte sind auch ganz aktuell wieder im Fokus der deutschen Kartellrechtsdebatte, nachdem das BMWK im Juni 2022 angesichts der hohen Benzin- und Dieselpreise neue Regelungen zu einer missbrauchsunabhängigen Entflechtung und zu einer Beweislasterkehr bei der Abschöpfung von kartellrechtswidrig erlangten Vorteilen ins Spiel gebracht hat.

Machtkonzentrationen sind auch in der digitalen Ökonomie nichts Neues. Schon als das Internet – das gerade einmal 30 Jahre alt ist – noch in den Kinderschuhen steckte, machte man sich Sorgen um die Marktmacht von Microsoft. In den späten 1990er Jahren wurde befürchtet, dass Microsoft seine Macht auf dem Markt für Betriebssysteme ausnutzen könnte, um auch andere Märkte – etwa für Webbrowser, Media Player und Medienstreaming – und

* Der Beitrag basiert auf einer am 15.6.2022 in Köln gehaltenen Vorlesung. Die Vortragssform wurde weitgehend beibehalten. Die Angabe zum DMA wurde an die am 12.10.2022 veröffentlichte deutsche Finalfassung angepasst.

letztlich das ganze Internet zu dominieren. Und als am 10. Januar 2000 der Medienkonzern TimeWarner und der Internet-Zugangspvoder AOL ihre Fusionspläne bekannt gaben, befürchteten manche, dass der neue Konzern das ganze Internet und damit auch wesentliche Meinungsmärkte unter seine Kontrolle bringen könnte. Beide Sorgen haben sich – wie wir heute wissen – als unbegründet erwiesen. Microsoft ist zwar (nach Apple) immer noch das zweitwertvollste Digitalunternehmen.¹ Aber das Internet dominieren heute andere und im Bereich der Mediaplayer und des Medienstreamings ist Microsoft bedeutungslos. Den Ton geben andere wie YouTube, Netflix oder Spotify an. Auch der Zusammenschluss TimeWarner/AOL erwies sich als Flop. Er bescherte TimeWarner schon 2002 rund 45 Mrd. U\$ Verlust.² 2009 wurde AOL ausgegliedert und 2015 für 4,4 Mrd. U\$ verkauft.³ AOL ist nicht – wie befürchtet – der Internetmonopolist geworden, sondern in die Bedeutungslosigkeit versunken.

Heute ist die Sorge, dass einzelne Unternehmen die digitale Wirtschaft, das Internet und die Medien dominieren könnten, auf Google, Amazon, Facebook und Apple, teils auch auf Microsoft, kurz: GAFA bzw. GAFAM gerichtet. Dass Google – gegründet 1997 – seine Führungsrolle auf den Suchmaschinenmärkten als Hebel nutzen könnte, um andere Märkte für Karten (Maps), Videoinhalte (YouTube) oder Preisvergleiche (Shopping) zu erobern, ist seit mehr als zehn Jahren Gegenstand heftiger Diskussionen. Hinzu kommt, dass Google (bzw. Alphabet) mit Android das erfolgreichste mobile Betriebssystem kontrolliert.⁴ Amazon hat sich seit seiner Gründung als Internetbuchhändler 1994 zum weltweit größten Internetmarktplatz entwickelt und betreibt nebenbei mit dem Amazon Marketplace, für den es die Regeln setzt, auch eines der größten Vertriebsportale für Drittanbieter. Facebook (bzw. Meta) – mit dem Geburtsjahr 2004 der Benjamin in diesem exklusiven Club – ist das weltweit größte soziale Netzwerk. Hier bestehen in besonderem Maße neben der Angst vor einer wirtschaftlichen Machtkonzentration auch Bedenken mit Blick auf die potentielle „Medienmacht“. Dies

1 <https://de.fxssi.com/groessten-technologieunternehmen>.

2 <https://www.manager-magazin.de/unternehmen/karriere/a-232896.html>.

3 Gekauft hatte TimeWarner AOL für 182 Mrd. U\$, s. <https://www.mediavillage.com/article/historys-moment-in-media-aol-time-warner-merger/>.

4 Vgl. <https://de.statista.com/statistik/daten/studie/184335/umfrage/marktanteil-der-mobil-betriebssysteme-weltweit-seit-2009/>.

gilt umso mehr, seit Facebook mit WhatsApp⁵ und Instagram⁶ zwei weitere wesentliche Kommunikationskanäle übernommen hat.

Wie schon in den Microsoft-Fällen treten bei diesen Diensten starke Netzwerk- und Skaleneffekte auf, also „Selbstverstärkungseffekte“ und andere Faktoren, die nicht nur, aber besonders für digitale Plattformen relevant sind. Dies im Detail zu erörtern, würde den Rahmen dieser Vorlesung sprengen. Das deutsche Recht hat aber die wichtigsten, insbesondere (aber nicht nur) für digitale Märkte relevanten Faktoren in § 18 Abs. 3a GWB aufgezählt. Hier sollen nur zwei exemplarisch herausgegriffen werden: Einerseits Netzwerkeffekte und Größenvorteile (§ 18 Abs. 3a Nr. 1 und 3 GWB) und andererseits der Zugang zu wettbewerbsrelevanten Daten (§ 18 Abs. 3 Nr. 3 und Abs. 3a Nr. 4 GWB).

Ein Betriebssystem, ein soziales Netzwerk oder ein Messenger-Dienst funktionieren einerseits umso besser, je mehr Teilnehmer der Dienst hat, mit denen man sich austauschen kann. Andererseits werden sie mit steigender Nutzerzahl auch für Werbekunden immer attraktiver. Man spricht insoweit von „Netzwerkeffekten“. Wer schon einmal erwogen hat, von WhatsApp zu einem anderen Dienst zu wechseln, weiß, wovon die Rede ist. Das macht es neuen Unternehmen schwer, auf einem Markt Fuß zu fassen. Sie müssen nicht nur erhebliche Investitionen für Hardware, Server und Programme, aber auch für die Gewinnung von Kunden tätigen, sondern zudem sowohl auf dem Nutzermarkt (z.B. für ein neues soziales Netzwerk) als auch auf dem Werbemarkt eine hinreichende Zahl von Kunden gewinnen und diese letztlich überzeugen, von dem bereits etablierten Dienst, den bereits „jeder“ nutzt, zu wechseln. Unmöglich ist das aber nicht. Google war nicht immer die führende Suchmaschine und Facebook nicht immer das führende soziale Netzwerk. Beide haben vormals dominante Wettbewerber „entthront“. Und nach der Übernahme von Skype hatte Microsoft 2011 kurzzeitig einen Marktanteil von über 90 % bei den Messenger-Diensten.⁷ Genützt hat es nichts. Heute dominieren andere Dienste wie WhatsApp oder Zoom, die – allen Netzwerkeffekten zum Trotz – Erfolg hatten, weil sie schlicht besser waren.

In den letzten Jahren hat sich der Blick zunehmend auf eine mutmaßliche „Datenmacht“ der GAFAM-Unternehmen gerichtet. Daten spielen für jede Industrie eine wichtige Rolle, für die Internetdienste von Google und Facebook gilt dies aber in besonderem Maße, weil diese Unternehmen den

5 KOMM., 19.11.2014, M.8228 – Facebook/WhatsApp.

6 Vgl. <https://about.fb.com/news/2012/04/facebook-to-acquire-instagram/>.

7 KOMM., 7.10.2011, M.6281 – Microsoft/Skype.

Nutzern ihre Dienste kostenfrei anbieten und sich über ein mehrseitiges Geschäftsmodell über Werbung finanzieren. Die Nutzerdaten werden genutzt, um die Produkte zu optimieren und zielgenauere und damit erfolgreichere Werbung zu schalten. Mit anderen Worten: Die Nutzerdaten sind ein wichtiger Rohstoff, den die Algorithmen dieser Unternehmen – also letztlich die KI, die Produkte und Werbung steuert – benötigen, um gut zu funktionieren. Aber anders als andere Rohstoffe oder Geld sind Daten „nicht rival“: Geld, das man ausgibt, ist weg. Öl, das man verbrennt oder verarbeitet, ist weg. D.h. Geld oder Öl kann nur der eine oder die andere nutzen. Es besteht Rivalität um diese Verbrauchsgüter. Bei Daten ist das oft anders. Die meisten Daten kann man mehrfach an verschiedene Unternehmen übermitteln, die sie alle parallel oder nacheinander nutzen können. Um mitzuhalten braucht man auch nicht unbedingt genauso viele Daten – oder gar die gleichen Daten – wie Google und Facebook. Aber man braucht jedenfalls „genug“ Daten, um seine KI zu trainieren. Man spricht insoweit von einer „mindestoptimalen Datenmenge“.⁸ Ob es wirklich „Datennetzwerkeffekte“ gibt, erscheint zweifelhaft.⁹

Mit Blick auf diese Faktoren und die schiere Größe der GAFAMs besteht heute gleichwohl die Sorge, dass diese Unternehmen bereits uneinholbare Wettbewerbsvorteile erlangt hätten und dass kleinere – insbesondere deutsche und europäische – Unternehmen ins Hintertreffen geraten, z.B. weil sie aus Datenmangel nicht mithalten können oder weil die GAFAMs sich selbst bevorzugen. Diese Sorge steht hinter dem neuen § 19a GWB und auch hinter dem Digital Markets Act (DMA) der EU (zu Deutsch: Gesetz über digitale Märkte). Außerdem steht die Sorge im Raum, dass innovative Start-ups durch Fusionen „geschluckt“ und dadurch Fortschritt und Produktvielfalt beschnitten werden. Am Ende, so die verbreitete Befürchtung, werden dann wir alle als Marktteilnehmer monetär oder in Bezug auf unsere Daten ausgebeutet.

8 Zu diesen Aspekten vgl. T. Körber, „Ist Wissen Marktmacht?“ Überlegungen zum Verhältnis von Datenschutz, „Datenmacht“ und Kartellrecht – Teil I, NZKart 2016, 303 und ders., „Ist Wissen Marktmacht?“ Überlegungen zum Verhältnis von Datenschutz, „Datenmacht“ und Kartellrecht – Teil II, NZKart 2016, 348 m.w.N.

9 Dazu und zu den vorstehenden Aspekten monographisch K. Hillmer, Daten als Rohstoffe und Entwicklungstreiber für selbstlernende Systeme, Baden-Baden 2021.

B. Kartellrecht

Aber wie dem auch sei: Wenn die Gefahr besteht, dass Unternehmen „zu groß“ sind und nicht mehr hinreichend durch Markt und Wettbewerb kontrolliert werden, schlägt – wie schon 1890 – zunächst einmal die Stunde des Kartellrechts. Da nicht alle Hörer der Ringvorlesung Kartellrechtsexperten sind, soll hier zunächst vorab kurz skizziert werden, was das Kartellrecht kann und will und auch, was es nicht kann und nicht will.

Das Kartellrecht dient dem Schutz der Freiheit des Wettbewerbs auf ökonomischen Märkten. Wirksamer Wettbewerb als freies Spiel der Marktkräfte sorgt dafür, dass das Angebot durch die Nachfrage gesteuert wird und dass knappe Ressourcen dort eingesetzt werden, wo sie volkswirtschaftlich ihre sinnvollste Verwendung finden. Die Geschichte lehrt, dass die *über* den Markt aktivierte „Schwarmintelligenz“ der (privaten wie unternehmerischen) Marktteilnehmer zu dieser Koordinierung weit besser in der Lage ist als staatliche Planung auf der Basis von meistens nur angemäßigem Wissen. Der Markt ist ein „basisdemokratischer“ Steuerungsmechanismus. Der Wettbewerb ist zugleich ein Instrument der Freiheitssicherung gegenüber großen Unternehmen wie gegenüber dem (noch größeren) Staat.

Wenn der Wettbewerb funktioniert, führt er zu niedrigen Preisen, Produktvielfalt und Innovation – kurz: zur Steigerung der Konsumentenwohl-*fahrt*. Wirksamer Wettbewerb setzt die Unternehmen beständig unter Druck, ihre Preise zu senken und ihre Produkte und Verfahrensweisen zu verbessern, weil sie sonst Marktanteile verlieren und ggf. sogar vom Markt verschwinden. Wie in der Natur gilt der Grundsatz des „survival of the fittest“. Dass Unternehmen, die innovativ und effizient sind, wachsen – und andere, die es nicht sind, „pleitegehen“ – ist Ausdruck des *Funktionierens* von Märkten und nicht etwa von Marktversagen.

Um die Funktionsfähigkeit und Wirksamkeit des Wettbewerbs zu sichern, verbietet das Kartellrecht *erstens* – wie schon der Name vermuten lässt – Kartelle, also Verhaltensabstimmungen zwischen Unternehmen, die Wettbewerbsbeschränkungen bezwecken oder bewirken, z.B. Preisabsprachen.¹⁰

Zweitens versucht das Kartellrecht, Machtkonzentrationen durch wettbewerbswidrige Unternehmenszusammenschlüsse zu verhindern. Dies ist Gegenstand der Fusionskontrolle.¹¹ Wenn also beispielsweise ein GAFAM-Unternehmen sich durch ein innovatives Start-up-Unternehmen bedroht

10 Z.B. § 1 GWB, Art. 101 AEUV.

11 Z.B. §§ 35 ff. GWB, EU-FKVO 139/2004.

fühlt und dieses „schluckt“, um sich einen Wettbewerber „vom Hals zu schaffen“, kann das ggf. verboten werden. Der Teufel steckt dabei freilich im Detail, denn eine solche Fusion kann auch innovationsfördernd wirken und seitens des Start-ups durchaus erwünscht sein.¹² Wer träumt nicht davon, dass ein wenige Jahre altes Start-up für Hunderte von Millionen oder sogar Milliarden aufgekauft wird, wie seinerzeit WhatsApp? Diese Aussicht kann ein starker Anreiz sowohl für Innovationen als auch für Investitionen in solche Innovationen sein und daher den Fortschritt antreiben. Fusionen werden daher in der Praxis zu Recht nur äußerst selten verboten.¹³

Drittens, und damit sind wir beim eigentlichen Thema, sollen die Machtmissbrauchsverbote des Kartellrechts verhindern, dass marktbeherrschende Unternehmen, die nicht mehr hinreichend durch den Wettbewerb kontrolliert werden, ihre Macht missbrauchen, um ihre Wettbewerber zu behindern oder ihre Marktpartner auszubeuten. Aber auch hierbei muss sorgfältig unterschieden werden, denn Erfolg im Wettbewerb führt – wie wir gesehen haben – dazu, dass einige Unternehmen größer werden und andere Marktanteile verlieren. Das Kartellrecht verbietet daher aus gutem Grund nicht, dass Marktmacht im Wettbewerb errungen wird. Das nämlich würde bedeuten, Erfolg zu bestrafen. Es kann geschehen, dass ein Unternehmen eine marktbeherrschende Stellung erlangt, ganz einfach, weil es innovativer, effizienter und schlicht „besser“ ist als seine Wettbewerber. Marktmacht an sich und sogar Monopole sind daher zu Recht erlaubt. Ein solches Unternehmen zu zerschlagen, auch wenn es seine Macht nicht missbraucht, wie in Bezug auf die GAFAMs und jetzt seitens des BMWK auch in Bezug auf Mineralölunternehmen immer wieder mal gefordert wird, ist daher ein kartellrechtsfremder Irrweg. Solche Entflechtungen wurden z.B. in den USA und im Vereinigten Königreich versucht und haben praktisch nie zu dauerhaften Erfolgen geführt. Das Kartellrecht kann und sollte nicht gegen den Markt anarbeiten, sondern nur dem freien Wettbewerb auf Märkten „Leitplanken“ setzen. Zielführender ist es daher, marktmächtigen Unternehmen kartellrechtlich „auf die Finger“ zu schauen und – eben weil sie nicht mehr hinreichend durch den Wettbewerb kontrolliert werden – Machtmissbräuche durch solche Unternehmen zu

12 Kritisch zum Vorwurf wettbewerbsbehindernder Akquisitionen T. Gerpott, Empirische Befunde zu Unternehmenskäufen durch Big Tech-Konzerne – Forschungssichtung und Implikationen für die Wettbewerbsrechtsanwendung und -setzung, NZKart 2022, 561.

13 Die Kommission hat von den bis September 2022 bei ihr angemeldeten 8654 Zusammenschlüssen nur 32 vollständig untersagt, vgl. <https://tlp.de/d5vsd>.

identifizieren, zu unterbinden und zu beußen. Erst wenn das nicht hilft, ist bei wiederholten Missbräuchen als ultima ratio gegenüber „kartellrechtlichen Intensivtätern“ schon nach geltendem Recht eine Entflechtung möglich.¹⁴

C. „Analoges Kartellrecht“ und digitale Märkte

Das Kartellrecht baut auf Generalklauseln. Dies gilt für das Kartellverbot (§ 1 GWB, Art. 101 AEUV) ebenso wie für das Missbrauchsverbot (§§ 19, 20 GWB, Art. 102 AEUV). Und das ist auch gut so, denn spezielle – am technischen Status quo ausgerichtete Normen würden von der Praxis schnell überholt sein. Das gilt sogar in besonderem Maße für digitale Märkte, weil sie außerordentlich dynamisch sind. Ein spezielles „Digital-Kartellrecht“ würde daher sogar besonders schnell veralten und überholt sein. Generalklauseln können dagegen von Kartellbehörden und -gerichten flexibel an die Entwicklungen der Praxis angepasst werden.¹⁵

Die kartellrechtlichen Missbrauchsverbote haben allerdings einen „Pferdefuß“: Ihre Durchsetzung durch Behörden oder Gerichte setzt einen bereits erfolgten Machtmissbrauch voraus. Sie können also – in anderen Worten – nur *ex post* angewendet werden. Manche meinen daher, das bestehende Kartellrecht leiste „too little, too late“ und könne nicht mit der Digitalwirtschaft mithalten.¹⁶ Wenn das Kartellrecht endlich eingreife, sei – z.B. im Verhältnis zu den GAFAMs – das Kind „Wettbewerb“ möglicherweise schon in den Brunnen gefallen. Paradebeispiel der Vertreter dieser These ist das *Google Shopping*-Verfahren: Hier vergingen zwischen der Beschwerde des Wettbewerbers *Foundem* am 3. November 2009 und der Entscheidung der EU-Kommission am 27. Juni 2017 nahezu acht Jahre,¹⁷ und final entschieden ist der Fall immer noch nicht. Vor dem Gericht der Europäischen Union hat die EU-Kommission am 10. November 2021 gewonnen, aber das letzte Wort hat der EuGH. Das Bundeskartellamt hat allerdings kürzlich ein Verfahren gegen *Amazon* in rund 9 Monaten abgeschlossen.¹⁸ Es geht also auch deutlich schneller. Warum dauerte es im ersten Verfahren so lange und ging im zweiten so schnell, obwohl beide auf der Basis des „alten“ Kartellrechts und beide

14 Vgl. § 32 Abs. 2 GWB, Art. 7 Abs. 1 VO 1/2003.

15 Vgl. dazu schon T. Körber, Analoges Kartellrecht für digitale Märkte?, WuW 2015, 120.

16 Vgl. etwa R. Podszun/P. Bongartz/S. Langenstein, The Digital Markets Act: Moving from Competition Law to Regulation for Large Gatekeepers, EuCML 2021, 60.

17 KOMM., 27.6.2017, AT.39740 – Google Shopping.

18 S. <https://t1p.de/2lujh>.

gegen Digitalunternehmen geführt wurden? Ein Grund mag sein, dass EU-Kommission und Bundeskartellamt im Jahr 2022 viel mehr Kompetenz im Bereich der digitalen Ökonomie haben als dies 2009 der Fall war. Wichtiger dürfte aber sein, dass im *Google Shopping*-Verfahren eine einvernehmliche Einigung zwischen Google und der EU-Kommission (zu der beiden bereit waren) durch politische Intervention verhindert und eine im Streit ergangene Entscheidung erzwungen wurde, während Amazon und das Bundeskartellamt ohne politische Intervention eine einvernehmliche Lösung suchen und finden konnten, die dann natürlich auch kein langwieriges Gerichtsverfahren nach sich zog. Dies zeigt, wie wichtig politische Unabhängigkeit für den zügigen Wettbewerbsschutz ist.

Gleichwohl ist richtig, dass die Anwendung der Missbrauchsvorschriften sehr aufwendig ist. Sie erfordert umfassende ökonomische und juristische Analysen. Und auch der Nachweis von Missbräuchen ist oft schwierig. Das gilt auch und vielleicht sogar besonders für digitale Märkte, die oft noch komplexer sind. Während sich auf einem „normalen“ Markt Anbieter und Nachfragende gegenüberstehen (z.B. Autohändler und Käufer), sind auf digitalen Märkten oft mehrere Marktseiten über Plattformen vernetzt. Bei Googles mobilem Betriebssystem Android sind es z.B. fünf Marktseiten, die sich gegenseitig beeinflussen. Mehrere 2019 und 2020 erstellte wissenschaftliche Gutachten für BMWi und EU-Kommission kamen trotzdem (zu Recht) zu dem Schluss, dass eine maßvolle Anpassung, digitale Ertüchtigung und Beschleunigung des bestehenden Kartellrechts ausreichen würde.¹⁹ Aber das Bundeskartellamt und die EU-Kommission strebten nach einer Erweiterung ihrer Kompetenzen und die Politik strebte nach einer stärkeren Intervention gegen die US-amerikanischen Unternehmen, um die Wettbewerbschancen deutscher und europäischer Unternehmen zu verbessern. Daher wurden in Deutschland und der EU Gesetze entwickelt, durch die man große Digitalunternehmen oder auch „Torwächter“ (Gatekeeper) vorab – *ex ante* – regulieren möchte. In Deutschland wurde eine solche Regel in § 19a GWB mit der 10. GWB-Novelle bereits am 19. Januar 2021 Gesetz. Auf Ebene der EU wurde am 12. Oktober 2022 der finale Text des Gesetzes über digitale Märkte, besser bekannt als Digital Markets Act – kurz DMA – im Amtsblatt veröffentlicht,

19 H. Schweitzer/J. Haucap/ W. Kerber/R. Welker, Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen, abrufbar unter: <https://tlp.de/x4pp>; BMWi, Wettbewerbskommission 4.0, Ein neuer Wettbewerbsrahmen für die Digitalwirtschaft, Bericht der Kommission Wettbewerbsrecht 4.0, abrufbar unter: <https://tlp.de/2owqj>; J. Crémer/Y.-A. de Montjoye/H. Schweitzer, Competition Policy for the Digital Era, abrufbar unter: <https://tlp.de/bnfa>.

der ab 2. Mai 2023 gilt. Auf diese beiden Regelwerke wirft der letzte Teil dieses Beitrags einen – angesichts der Komplexität der Bestimmungen notwendig summarischen und auch kritischen – vergleichenden Blick.

D. Exkurs: Es muss nicht immer Kartellrecht sein!

Zuvor ist es aber wichtig zu unterstreichen, dass das Kartellrecht kein „Allheilmittel“ ist, um die Probleme digitaler (oder anderer) Märkte zu lösen. Oft ist es dafür nicht oder jedenfalls schlechter geeignet als andere Gesetze. Das zeigt der laufende *Facebook*-Fall des Bundeskartellamtes, bei dem es primär um mutmaßlich unangemessene AGB und Datenschutzverletzungen geht.²⁰ Hier stellt sich die Frage, ob nicht in erster Linie das AGB-Recht und das Datenschutzrecht anzuwenden sind. Das gilt umso mehr, als die Bürgerinnen und Bürger gegenüber *jedem* Unternehmen vor dem Missbrauch von AGB und Datenschutzverstößen geschützt werden und nicht nur gegenüber Torwächtern. Dass hier und auch in anderen Fällen dennoch häufig zum Kartellrecht gegriffen wird, liegt einerseits daran, dass hier schlagkräftige Behörden existieren, während die Durchsetzung anderer Normen (z.B. des AGB-Rechts) nur über Klagen erfolgt. Andererseits können im Kartellrecht sehr hohe Bußgelder verhängt werden, in den drei *Google*-Verfahren der EU-Kommission waren es zusammen sage und schreibe 8,25 Mrd. Euro.²¹

Die fortlaufende Anpassung des Kartellrechts ist daher nur ein Puzzleteil in einem wahren Gewirr neuer Regelungen im Bereich der Digitalisierung. Die EU hat beispielsweise eine Vielzahl von Regelungen erlassen oder „in der Pipeline“, die allesamt mehr oder weniger digitale Märkte betreffen. Um nur die wichtigsten zu nennen: 2018 wurde die Datenschutz-Grundverordnung zum Schutz personenbezogener Daten erlassen, die z.B. in ihrem Art. 20 ein Recht auf Datenportabilität enthält. Ebenfalls 2018 erließ die EU die Verordnung 2018/1807 über den grenzüberschreitenden Fluss nicht-personenbezogener Daten. 2018 wurde auch die TK-Kodex-Richtlinie 2018/1972 erlassen, die z.B. Regelungen zu Messenger-Programmen und zur Interoperabilität enthält und gerade im neuen TKG 2021 umgesetzt wurde. 2022 folgte der Data Governance Act (VO 2022/868), der das grenzüberschreitende Teilen von

20 S. BKartA, Fallbericht v. 15.2.2019 – B6–22/16; dazu kritisch *T. Körber*, Die Facebook-Entscheidung des Bundeskartellamtes – Machtmissbrauch durch Verletzung des Datenschutzrechts?, NZKart 2019, 187 m.w.N.

21 S. Kommissionsentscheidungen *Google Shopping* (AT.39740), *Google Search/Ad Sense* (AT.40411) und *Google Android* (AT.40099).

Daten erleichtern und Intermediäre regulieren soll, ebenso der Digital Services Act (Gesetz über digitale Dienste, VO 2022/2065, DSA), der – gestaffelt nach der Größe der Plattformen – inhaltsbezogene Regeln für die Online-Plattformen aufstellt, die z.B. gegen sog. „hate speech“ oder Nutzerirreführung durch sog. „dark patterns“ gerichtet sind, und ab 17. Februar 2024 gelten wird. Vor allem aber wurde 2022 auch der bereits erwähnte Digital Markets Act (DMA) finalisiert, der die wirtschaftliche Macht von Internet-Gatekeepern begrenzen und auf diese Weise – ähnlich zum Kartellrecht – die Bestreitbarkeit von Märkten und die Fairness in Märkten sicherstellen will, auch wenn er in seinen Erwägungsgründen 10 und 11 betont, kein Kartellrecht zu sein. In der Gesetzgebungs-Pipeline sind u.a. die ePrivacy-Verordnung über das Datensammeln im Internet, die an die Stelle der bisherigen „Cookie-Richtlinie“ 2002/58 treten soll. Sie sollte eigentlich schon 2018 mit der DSGVO verabschiedet werden, hat sich aber immer wieder verzögert. Gleiches gilt für den Data Act, der in allgemeiner Weise Regeln zum Datenzugang und zur Datennutzung in der EU setzen soll, um Interoperabilität sicherzustellen und Datenhandel zu ermöglichen, und für den Artificial Intelligence Act, der den Einsatz von Künstlicher Intelligenz regeln und zum Schutz der Bürgerinnen und Bürger begrenzen will, namentlich in Bezug auf Produktsicherheit, Diskriminierung und Datensicherheit.

Diese Regelungen sind nicht immer miteinander oder gar mit den parallel laufenden nationalen Regelungen abgestimmt. Für große Unternehmen ist das sicher lästig und teuer, aber handhabbar. Und für die Anwältinnen und Anwälte ist das sogar gut, weil sich ein El Dorado künftiger Beratungsbedarfe und Rechtsstreitigkeiten auftut.²² Aber kleine und mittlere Unternehmen und insbesondere Start-ups können dadurch möglicherweise vom Marktzutritt abgeschreckt oder aus dem Markt oder in Fusionen gedrängt werden. Kurzgefasst: Zu viel Regulierung kann, obwohl sie Wettbewerb ermöglichen will, am Ende dazu führen, dass es zu einer stärkeren Unternehmenskonzentration und damit zu weniger Wettbewerb kommt.

22 Vgl. zur Ressourcenbindung durch Verfahren auf Grundlage des DMA I. *Brinker/K. Haag*, Regulierung (mehrseitiger) Online-Plattformen: Neue Ansätze im Kartellrecht, Juristentag diskutiert über eine verstärkte Regulierung von Digitalunternehmen, AnwBl-Onl. 2022, 485 (489).

E. Torwächter-Regulierung nach § 19a GWB und DMA

Die neuen Regelungen in § 19a GWB und im DMA wollen den beschriebenen Problemen der mutmaßlichen Langsamkeit und Komplexität der *ex post*-Anwendung des Kartellrechts abhelfen, indem sie einerseits bestimmten „großen Digitalunternehmen“ Vorgaben machen, die nicht erst greifen, wenn ihnen ein Missbrauch nachgewiesen wurde, sondern diese Unternehmen vorab – *ex ante* – „an die Kette legen“, und andererseits den Kartellbehörden das Leben durch Beweiserleichterungen leichter und auch dadurch die Rechtsdurchsetzung schneller machen. Zugleich wollen sie ein „level playing field“ schaffen, um auf diese Weise die Wettbewerbschancen kleinerer – vor allem auch deutscher und europäischer – Wettbewerber zu verbessern. Auf diese Weise sollen Wettbewerb auf Märkten und Fairness in Märkten gefördert werden.²³ Dagegen ist im Grundsatz nichts zu sagen. Der Teufel steckt aber im Detail.

I. Torwächter-Bestimmung nach § 19a Abs. 1 GWB und Art. 3 DMA

Zunächst muss bestimmt werden, wer überhaupt Adressat der neuen *ex ante*-Regelungen sein soll. Beide Gesetze machen dafür bestimmte Vorgaben und beide erfordern eine Behördenentscheidung. Beide sehen vor, dass Unternehmen via Verwaltungsakt zu Normadressaten erklärt werden. Das ist richtig, denn die Unternehmen brauchen Rechtssicherheit. Die Kriterien dafür sind im deutschen Recht und im DMA der EU allerdings recht unterschiedlich.

Schon die jeweils verwendeten Begriffe weichen ab. § 19a Abs. 1 spricht von „Unternehmen mit überragender marktübergreifender Bedeutung für den Wettbewerb“, kurz ÜMB-Unternehmen oder mit dem BKartA „ÜmüB“. Man könnte insoweit auch von „digitalen Torwächtern“ sprechen (wie es der DMA tut), denn gemeint sind ausweislich der Gesetzesbegründung „große Digitalkonzerne“²⁴ oder – um das Kind beim Namen zu nennen – die US-amerikanischen GAFAM-Unternehmen. Es geht in der Sache also um die Regulierung der großen digitalen Internetplattformen. Um die Frage nach dem „Wer“ zu beantworten, zählt § 19a GWB eine Reihe von qualitativen Faktoren

23 Vgl. RegBegr zur 10. GWB-Novelle, BT-Drucks. 19/23492, S. 56 und 73; Erwägungsgründe 1 ff. zum DMA.

24 RegBegr, BT-Drucks. 19/23492, S. 73.

als Indizien auf, die auf eine „überragende marktübergreifende Bedeutung“ hindeuten können, darunter Marktbeherrschung, Finanzkraft, Integration, Datenzugang und Plattformmacht. Das Bundeskartellamt muss diese Faktoren untersuchen. Am Ende trifft es eine Ermessenentscheidung und erlässt einen Verwaltungsakt, mit dem es ein Unternehmen für zunächst fünf Jahre zum ÜMB-Unternehmen erklärt. Wenig überraschend hat das Bundeskartellamt im ersten Jahr nach Inkrafttreten der Norm vier Verfahren – gegen Google, Amazon, Facebook und Apple – eingeleitet. Google wurde bereits im Dezember 2021, Facebook (Meta) im Mai 2022 und Amazon im Juli 2022 als ÜMB-Unternehmen eingestuft. Ein Verfahren gegen Apple läuft noch.²⁵

Auch nach dem DMA legt eine Behörde, hier die EU-Kommission, fest, wer ein Torwächter ist. Diese Entscheidung gilt für drei Jahre. Anders als § 19a GWB ist der DMA dabei auf bestimmte Internetdienste, sog. „zentrale Plattformdienste“ (core platform services) beschränkt, die in seinem Art. 2 aufgelistet werden. In Art. 2 findet sich ein bunter Strauß von Diensten, namentlich Online-Vermittlungsdienste, Online-Suchmaschinen, soziale Netzwerke, Video-Sharing-Plattformen, Messenger-Dienste, Betriebssysteme, Webbrowser, virtuelle Assistenten, Cloud-Computing und Online-Werbung. Etwas salopp könnte man formulieren: „Fast alles, was die GAFAMs so machen“. Wird ein solcher zentraler Plattformdienst erbracht, muss die EU-Kommission – anders als das BKartA – nicht unbedingt eine umfassende Abwägung *qualitativer* Kriterien hinsichtlich einer marktübergreifenden Bedeutung vornehmen. Sie kann sich auf Art. 3 Abs. 2 DMA verlassen, der die Torwächter-Eigenschaft vermutet, wenn ein Unternehmen bestimmte quantitative Schwellenwerte überschreitet, nämlich unionsweite Umsätze von mindestens 7,5 Mrd. Euro über drei Jahre oder eine Marktkapitalisierung von mindestens 75 Mrd. Euro (Art. 3 Abs. 2 lit. a) und mindestens 45 Mio. monatlich aktive Endnutzer und 10.000 jährlich aktive Geschäftskunden in der EU in den letzten drei Jahren (Art. 3 Abs. 2 lit. b und c). Die mutmaßlichen Torwächter können zwar theoretisch nachweisen, dass sie trotz Erfüllung dieser Schwellenwerte keine Torwächter sind, aber die Hürden dafür sind so hoch, dass das in der Praxis kaum gelingen dürfte.

Auf den ersten Blick liefert der DMA durch die Auflistung der betroffenen Dienste und die klaren quantitativen Schwellenwerte mehr Rechtssicherheit als das GWB. Insbesondere kleine Dienstleister können leichter erkennen,

25 Vgl. BKartA, 30.12.2021, Az. B7–61/21 – Google: Feststellung der ÜMB; 2.5.2022, Az. B6–27/21 – Meta: Feststellung der ÜMB; 5.7.2022, Az. B2–55/21 – Amazon: Feststellung der ÜMB.

dass sie nicht unter den DMA fallen. Doch sind die Maßstäbe allzu grob geraten. *Erstens* verwechselt der DMA Unternehmensgröße und „Gatekeeping“. Wenn ein Unternehmen hinreichend groß ist, wird es zum Torwächter erklärt, ohne auch nur das Tor zu nennen, das es angeblich bewacht. *Zweitens* sind die Schwellenwerte für alle Dienste gleich, obwohl 45 Mio. monatliche Nutzer für eine Suchmaschine so gut wie nichts, für andere Dienste dagegen möglicherweise sehr viel sind. Und *drittens* basieren diese Schwellenwerte nicht auf irgendwelchen ökonomisch fundierten Fakten. Sie sind das Resultat eines rein politisch ausgehandelten Deals und wurden im Vergleich zum DMA-Entwurf von Dezember 2020 nochmal hochgesetzt, wohl um möglichst wenige europäische Unternehmen zu erfassen.

II. Torwächter-Regulierung nach § 19a Abs. 2 GWB und Art. 5, 6 und 7 DMA

Was geschieht, wenn ein Unternehmen durch einen ersten Verwaltungsakt zum Torwächter erklärt wurde? Auch hier zeigen beide Regelungen Gemeinsamkeiten und Unterschiede. Während sich die Verbotskataloge ähneln, ist das Durchsetzungsverfahren unterschiedlich gestaltet.

Anders als das Kartellrecht, das mit flexiblen Generalklauseln arbeitet, enthalten § 19a Abs. 2 GWB und der DMA umfassende und teils sehr detaillierte Kataloge von Ge- und Verboten. Bei § 19a Abs. 2 GWB sind es sieben, die teils durch Varianten und Regelbeispiele weiter ausdifferenziert werden: Selbstbevorzugung (Nr. 1) und die Behinderung von Wettbewerbern auf Zugangs- und Beschaffungsmärkten (Nr. 2), das „Aufrollen“ von Märkten, d.h. der Eroberung eines Marktes durch ein schnelles internes Wachstum aufgrund überragender Ressourcen (Nr. 3), die Behinderung von Wettbewerbern durch Datennutzung (Nr. 4), die Verweigerung oder Erschwerung von Interoperabilität und Datenportabilität (Nr. 5), die unzureichende Information anderer Unternehmen über eigene Leistungen (Nr. 6) und das „Anzapfen“ von Marktpartnern, d.h. das Fordern unangemessener Vorteile (z.B. Rabatte oder Datennutzungsrechte) (Nr. 7).²⁶ Der DMA wartet in Art. 5, 6 und 7 sogar mit rund 30 sehr komplex und verschachtelt formulierten Ge- und Verboten auf, die aber in eine ähnliche Richtung gehen und hier nicht im Einzelnen

26 Dazu im Einzelnen (aus der Warte des BMWi) T. Käseberg, in: F. Bien/T. Käseberg/G. Klumpe/T. Körber/K. Ost, Die 10. GWB-Novelle, München 2021, Kap. 1 Rn. 193 ff. sowie (aus der Warte der Anwaltspraxis) M. Esser/C. Höft, in: F. Bien/T. Käseberg/G. Klumpe/T. Körber/K. Ost, Die 10. GWB-Novelle, München 2021, Kap. 1 Rn. 231 ff.

dargestellt werden können. Wenn man genauer hinschaut, wird deutlich, dass es sich bei diesen Katalogen um eine Art „best of“²⁷ oder „wünsch dir was“ der Kartellbehörden handelt, denn sie spiegeln die Auffassung von Bundeskartellamt bzw. EU-Kommission in – meistens noch laufenden – Kartellverfahren wieder. So stand beispielsweise für § 19a Abs. 2 Nr. 1 (Selbstbevorzugung) offensichtlich der *Google Shopping*-Fall der EU-Kommission Pate²⁸ und für Nr. 4 (Datennutzung) das *Facebook*-Verfahren des BKartA.²⁹ Ähnliche Ursprünge lassen sich für fast alle Ge- und Verbote aufzeigen, was mit Blick auf den DMA umso bemerkenswerter erscheint, weil dieser von sich behauptet, kein Kartellrecht sondern eine zum Kartellrecht lediglich komplementäre Regelung zu sein.³⁰

Die in den beiden Regelungen aufgelisteten Verhaltensweisen sollen Torwächtern in Zukunft verboten werden, ohne dass man ihnen vorher aufwändig einen Missbrauch ihrer Macht nachweisen muss. Das klingt erst einmal gut, ist aber alles andere als unproblematisch. Um nur ein paar Beispiele zu nennen: Das Verbot der Selbstbevorzugung entspricht *erstens* zwar der Auffassung der EU-Kommission im *Google Shopping*-Verfahren, doch hat der BGH auf anderen Gebieten eine solche Selbstbevorzugung in ständiger Rechtsprechung für wettbewerbskonform erklärt.³¹ Vor allem aber ist keineswegs immer klar, was überhaupt „Selbstbevorzugung“ ist. Ist es eine verbotene Selbstbevorzugung, dass ich meine Daten oder auch meine Server allein nutzen möchte? Ist es ein Ausdruck von Selbstbevorzugung, dass der Suchalgorithmus von Google die Ergebnisse anders reiht als der von Bing – oder ist das schlicht Suchmaschinenwettbewerb? Ähnliches gilt *zweitens* für den „Datenmissbrauch“: Behindere ich meine Wettbewerber, wenn ich legal erworbene Daten nutze und nicht mit ihnen teile? Und was wäre die Folge? Muss ich die Daten dann teilen? Oder darf ich sie selbst nicht mehr nutzen, auch wenn das bedeutet, dass meine Dienste für die Nutzer schlechter werden? Diese Unklarheiten bestehen *drittens* nicht minder mit Blick auf die Forderung nach Interoperabilität (z.B. von Messenger-Diensten wie WhatsApp): Ist es wirklich im Sinne der Nutzer, dass ein Messenger wie WhatsApp „interoperabel“ mit anderen wird, obwohl Multi-Homing möglich ist? Das ist wettbewerblich ambivalent, denn es kann einerseits dazu

27 Podszun/Bongartz/Langenstein, The Digital Markets Act (Fn. 16), 65.

28 KOMM., 27.6.2017, Fall AT.39740 – Google Shopping.

29 BKartA, 6.2.2019, Az. B6–22/16 – Facebook.

30 Vgl. Erwägungsgründe 10 und 11.

31 Z.B. BGH GRUR 2012, 84 (86) Rn. 31 – Telefon- und Branchenverzeichnisse; BGH NJW 2012, 2110 (2111) Rn. 15 – Freundschaftswerbung.

führen, dass die Wettbewerber sich den dafür notwendigen Vorgaben des Marktführers anpassen, statt diesem durch Innovationen Wettbewerb zu machen, und andererseits steigt dadurch die Gefahr für Spam, Hate-Mails und Sicherheitslücken, weil man sich insoweit möglicherweise auf den „kleinsten gemeinsamen Nenner“ einigen muss und – wenn Probleme auftauchen – ggf. nicht schnell reagieren kann. Das bedeutet nicht, dass in Bezug auf diese Aspekte keine Wettbewerbsprobleme bestehen, zeigt aber, dass man die reale Komplexität der digitalen Welt nicht einfach „wegdefinieren“ kann, indem man nur auf den ersten Blick klare Ge- und Verbote formuliert und die Torwächter „ohne Wenn und Aber“ an diese bindet.

Gerade in diesem Punkt – der Frage nach der Durchsetzung der Ge- und Verbote – unterscheiden sich § 19a GWB und der DMA maßgeblich. § 19a Abs. 2 GWB enthält keine unmittelbar kraft Gesetzes wirkenden Verbote. Die Norm ist lediglich eine Ermächtigungsgrundlage, aufgrund derer das Bundeskartellamt gegen ÜMB-Unternehmen Ge- oder Verbote im Wege eines weiteren Verwaltungsakts aussprechen kann. Dies gibt dem Bundeskartellamt die Möglichkeit, die Maßnahmen je nach Fallgestaltung sachgerecht anzupassen und eine Über- oder Unterregulierung zu vermeiden. Anders als beim klassischen Kartellrecht muss das Bundeskartellamt ÜMB-Unternehmen allerdings keinen Verstoß nachweisen. Die Verhaltensweisen gelten als missbräuchlich, wenn und weil ein ÜMB-Unternehmen sie an den Tag legt. Die Beweislast ist also umkehrt. Es gilt eine „gesetzliche Schuldvermutung“ und die ÜMB-Unternehmen müssen ihre Unschuld nachweisen. Verfassungsrechtlich ist schon diese Beweislastumkehr angesichts der oben beschriebenen wettbewerblichen Ambivalenz vieler Ge- und Verbote fraglich, weil es sich um staatliche Eingriffsverwaltung handelt und weil der Staat nach allgemeinen Regeln Eingriffe in Grundrechte rechtfertigen muss – und nicht umgekehrt Bürger oder Unternehmen deren Ausübung.³² Aber immerhin steht es den Unternehmen nach § 19a Abs. 2 S. 2 GWB frei, ihr Verhalten objektiv zu rechtfertigen.

Der DMA zeigt, dass man es noch viel schlechter machen kann als der deutsche Gesetzgeber bei § 19a GWB. Im Vergleich zum DMA ist § 19a GWB geradezu ein Musterbeispiel für Besonnenheit und Ausgewogenheit. Wie wir gesehen haben, wird beim DMA schon die Torwächter-Stellung mit einer Art

32 Vgl. auch *R. Polley/R. Kaup*, Paradigmenwechsel in der deutschen Missbrauchsaufsicht – Der Referentenentwurf zur 10. GWB-Novelle, NZKart 2020, 113 (118); *C. Degenhart*, Verfassungsfragen einer 10. GWB-Novelle auf der Grundlage des Referentenentwurfs vom 24.01.2020, WuW 2020, 308 (311 f.).

„Holzhammer“-Methode bestimmt, weil von der bloßen Größe auf die Torwächtereigenschaft geschlossen wird, ohne auch nur das bewachte Tor zu definieren. Diese Methode setzt der DMA leider auch bei der Durchsetzung der Ge- und Verbote fort. Diese sollen nämlich – damit es möglichst schnell geht – unmittelbar kraft Gesetzes gelten – also „self-executing“ sein, und gelten für alle Märkte und alle Torwächter nach einem „one size fits all“-Ansatz. Zu Deutsch: Alle Torwächter werden, ohne Rücksicht auf die unterschiedlichen Produkte, Geschäftsmodelle, Märkte und realen Wettbewerbssituationen „über einen Leisten“ geschlagen. Wurde ein Unternehmen zum Torwächter erklärt, so muss es binnen sechs Monaten alle Verpflichtungen nach Art. 5, 6 und 7 DMA erfüllen und dies nachweisen (Art. 3 Abs. 10 DMA). Die Ge- und Verbote nach Art. 6 und 7 DMA *kann* die EU-Kommission zwar noch genauer durch einen Verwaltungsakt ausgestalten, sie muss es aber nicht (Art. 8 DMA). Die Unternehmen haben auch keine Möglichkeit, ihr Verhalten objektiv zu rechtfertigen und z.B. nachzuweisen, dass die Nichterfüllung eines Ge- oder Verbots Wettbewerb, Innovation und die Konsumentenwohlfahrt gar nicht einschränkt, sondern vielleicht sogar fördert. Ein „Entlastungs- oder Unschuldswert“ ist gesetzlich ausgeschlossen.³³ Die Unternehmen müssen die Regeln befolgen, auch wenn dies nachweislich schlecht für den Wettbewerb und die Konsumentenwohlfahrt ist. Eine Aussetzung der Pflichten kommt nur ganz ausnahmsweise (Art. 9 DMA) und eine Befreiung davon nur zum Schutz der öffentlichen Gesundheit oder Sicherheit in Betracht (Art. 10 DMA). Kurzum: Schnelligkeit der Rechtsanwendung geht der Qualität der Rechtsanwendung und der Gewährleistung der Verfahrensgrundrechte vor. Das ist der Komplexität der Digitalwirtschaft nicht angemessen und wird in der Realität auch eher langsamer als schneller wirken als der deutsche Ansatz, weil es Rechtsstreitigkeiten provoziert, die sich dann – wie im *Google Shopping*-Verfahren – über viele Jahre hinziehen können.³⁴

33 Kritisch hierzu *D. Zimmer/J.-F. Göhsl*, Vom New Competition Tool zum Digital Markets Act: Die geplante EU-Regulierung für digitale Gatekeeper, *ZWeR* 2021, 29 (54 ff.); *H. Schweitzer*, The Art to Make Gatekeeper Positions Contestable and the Challenge to Know What Is Fair: A Discussion of the Digital Markets Act Proposal, *ZEuP* 2021, 503 (537 f.); *Körber*, Lessons from the Hare and the Tortoise: Legally imposed self-regulation, proportionality and the right to defence under the DMA – Part 2, *NZKart* 2021, 436 (439).

34 Kritisch hierzu etwa *Körber*, Lessons from the Hare and the Tortoise: Legally imposed self-regulation, proportionality and the right to defence under the DMA – Part 1, *NZKart* 2021, 379 und *ders.*, Lessons from the Hare and the Tortoise – Part 2 (Fn. 33), 436 m.w.N.

F. Fazit

Droht also der Untergang des digitalen Abendlandes, wenn der DMA 2023 Geltung erlangt? Sicher nicht! Die gute Nachricht für Sie, die Sie sich für Digitalisierung und Kartellrecht interessieren, ist, dass durch § 19a GWB, besonders aber durch den DMA Beratungsbedarf und Rechtsstreitigkeiten sicher einen neuen Aufschwung erleben werden. Kartellrechtsexpertinnen und -experten werden auch weiter ein sehr knappes und sehr teuer bezahltes Gut bleiben.

Wird der DMA den Wettbewerb beflügeln und werden digitale Champions in Deutschland und Europa aus dem Boden sprießen? Wohl auch nicht! Nationale oder europäische Champions kann man nicht „herbeiregulieren“ – und es ist sicher auch nicht der Mangel an Regulierung, der dazu geführt hat, dass solche Champions in Europa bisher nicht entstanden sind – eher im Gegenteil!

Eine Torwächter-Regulierung an sich kann durchaus Sinn ergeben, punktuell vielleicht sogar erforderlich sein. Aber sie sollte nicht mit dem Holzhammer erfolgen, sondern mit Augenmaß. Das deutsche Recht wird diesem Anspruch deutlich besser gerecht als der DMA.

Das Ganze erinnert an ein Bonmot, das die Mentalitätsunterschiede zwischen Preußen und Österreichern beschreibt und das – in einer seiner vielen Varianten – in den Befreiungskriegen gegen Napoleon verortet wird: Als es für die gegen Napoleon verbündeten Preußen und Österreicher auf dem Schlachtfeld schlecht aussieht, lässt der preußische General seinem Kollegen melden: „Die Lage ist ernst, aber nicht hoffnungslos“. Der Österreicher antwortet: „Die Lage ist hoffnungslos, aber nicht ernst“. Die EU-Kommission beschreibt die Lage in preußischer Manier als ernst, aber nicht hoffnungslos und versucht, durch den DMA gegenzusteuern. Näher liegt es, die Lage als „hoffnungslos, aber nicht ernst“ zu sehen: Es erscheint als vergleichsweise hoffnungsloses Unterfangen, die GAFAM-Unternehmen oder jedenfalls ihre Vormachtstellung „wegregulieren“ zu wollen, solange die Endnutzer ihre Produkte bevorzugen. Aber genau aus diesem Grunde ist die Lage auch nicht so ernst, wie mancher befürchtet, denn die mutmaßlichen „Gatekeeper“ haben sich oft auch als „Gateopener“ für neue Unternehmen wie Snapchat, TikTok oder Zoom erwiesen. Diese Newcomer konnten sich allen ökonomischen Glasperlenspielen, praktischen Widrigkeiten und politischen Schwarzmalereien zum Trotz durchsetzen, ganz einfach, weil sie besser sind

als die übermächtigen und scheinbar unangreifbaren „Platzhirsche“ und ihre Produkte. Genau darum geht es beim Wettbewerb. Und so soll es auch sein!

... denn sie wissen, was wir tun¹

Axel Ockenfels

Die meisten Menschen glauben, dass sie sich unter Kontrolle haben. Doch unsere Entscheidungen, Gefühle und Gedanken können immer besser prognostiziert und gesteuert werden. Die Vermessung des Menschen und die Berechnung seiner Zukunft bergen auf der einen Seite großartige Chancen, unser Leben erfolgreicher zu meistern. Sie erzeugen aber auch Befürchtungen vor Missbrauch und Ausbeutung. In diesem Beitrag skizziere ich ausgesuchte Beiträge der Erforschung von Rationalität und Künstlicher Intelligenz bei der Entschlüsselung des Menschen. Ich beschränke mich dabei auf das Verhalten in strategischen Situationen, in denen verschiedene Akteure Konflikte austragen, sich koordinieren müssen oder Kooperation anstreben.

A. Intelligenz, Rationalität und Prognose

Die wichtigste Voraussetzung für kluge Entscheidungen in unsicheren, strategischen Situationen ist die Fähigkeit, das Verhalten und die Reaktionen der anderen Spieler vorherzusagen. Eine gute Prognose ist der Schlüssel zum Erfolg in so unterschiedlichen Kontexten wie Schach oder Poker, auf Finanzmärkten, in der Politik, beim Management von Unternehmen, in Auktionen und auf Märkten, bei Verhandlungen und der Kindererziehung.

Doch wie können wir wissen, wie andere Akteure ticken, was sie als nächstes vorhaben und wie sie auf unsere Strategie reagieren werden? Zwei Prognosemodelle haben sich in vielen Situationen als erfolgreich erwiesen und ergänzen sich in wunderbarer Weise. Die *Spieltheorie* benötigt keine Daten, basiert auf der Annahme rationalen Verhaltens und erfreut sich großer Beliebtheit bei Wirtschaftswissenschaftlern. Die *künstliche Intelligenz* benö-

¹ Ich bedanke mich beim European Research Council (ERC, im Rahmen des European Union's Horizon 2020 research and innovation programme, GA No 741409 – EEC) und der Deutschen Forschungsgemeinschaft (DFG, im Rahmen der Exzellenzstrategie des Bundes und der Länder – EXC 2126/1- 390838866) für die Unterstützung meiner Forschung.

tigt dagegen sehr viele Daten, kommt dafür aber ohne Verhaltensannahmen aus und erfreut sich großer Beliebtheit in der Computerwissenschaft.

Obwohl die beiden Ansätze auf den ersten Blick nur wenig gemein haben, stehen beide Ansätze im Kern auf derselben Konzeption kluger Entscheidungen: Intelligent beziehungsweise rational ist, wer sich so verhält, dass das eigene Ziel aller Voraussicht nach bestmöglich erreicht werden kann.² Da der Schlüssel für eine erfolgreiche Zielerreichung eine gute Prognose ist, spielt die Frage, wie man Verhalten möglichst gut prognostizieren kann, in beiden Forschungsrichtungen die zentrale Rolle.

B. Rationalität und Superrationalität bei Schere – Stein – Papier

Schere – Stein – Papier ist das vielleicht einfachste Beispiel für eine strategische Situation. Jeder Spieler überlegt, was wohl der andere Spieler vorhat, und versucht, dem anderen einen Schritt voraus zu sein. Die bessere Prognose gewinnt.

Wie ist es möglich, in einer solchen Situation, ganz ohne die Hilfe von Daten, Verhalten vorherzusagen? Die Spieltheorie erklärt, wie man mit purer Logik und allein auf Basis von Rationalität zu einem Ergebnis kommt. Man erkennt schnell, dass rationale Spieler von Schere – Stein – Papier versuchen müssen, sich selbst unvorhersehbar zu machen. Daraus folgt, dass die Spieler in jeder Runde ganz unabhängig davon, was bisher passiert ist, jede der drei möglichen Optionen mit der gleichen Wahrscheinlichkeit auswählen. Jede Abweichung von dieser Strategie – jede systematische Bevorzugung einer der drei Optionen – könnte von dem Gegenüber ausgebeutet werden und so die Wahrscheinlichkeit erhöhen, dass man verliert. Bei Rationalität gibt uns die Spieltheorie so eine klare und eindeutige Vorhersage über das Verhalten der Spieler. Mit dieser Strategie kann jeder Spieler für sich garantieren, dass er keinen systematischen Nachteil besitzt und mit mindestens die gleiche Gewinnchance besitzt.

Die Sache wird zwar schnell kompliziert, aber theoretisch gibt es keine Grenzen für die Anwendbarkeit der Spieltheorie. Für (fast) alle strategischen Situationen können im Prinzip kluge Strategien berechnet werden und dadurch die Verhaltensweisen rationaler Akteure prognostiziert werden. Das ist nicht bloß theoretisch interessant; bei zahlreichen Anwendungen in der

2 S. J. Russell, *Human Compatible, Artificial Intelligence and the Problem of Control*, New York City, New York: Viking Press 2019.

Praxis lassen sich durch logische Schlussfolgerungen wichtige konzeptionelle Einsichten ableiten. So spielt das Rationalitätskonzept beispielsweise eine gewichtige Rolle bei der Verhaltensprognose in Spielen wie Poker, in Verhandlungen und Auktionen, bei Wettbewerbsstrategien unter Konkurrenten, bei Handelskriegen, politischen Wettkämpfen und militärischen Konflikten, beim Finanzmarkt- und Onlinehandel, bei Kooperationsfragen und bei der Verkehrssteuerung.

Die künstliche Intelligenz beschäftigt sich oft mit ähnlichen Herausforderungen, geht aber anders vor. Sie prognostiziert nicht auf Basis logischer und kausaler Schlussfolgerungen, sondern durch datenbasiertes Lernen. Hier gilt es vornehmlich, mithilfe maschinellen Lernens in großen Datensätzen statistische Zusammenhänge zu erkennen, die dann ihrerseits Vorhersagen erlauben. Autonomes Fahren prognostiziert zum Beispiel auf der Basis von sehr großen Datenmengen, wie sich menschliche Verkehrsteilnehmer verhalten würden.

Die Spieltheorie besitzt den Vorteil, dass sie auch Vorhersagen für Situationen treffen kann, für die es noch keine Daten gibt. Dafür kann datenbasierte Intelligenz ein Menschenmodell bauen und so oft bessere Strategien entwerfen als rationale Strategien im Sinne der Spieltheorie – man könnte diese Strategien als „superrational“ bezeichnen. Wenn beispielsweise große Datenmengen über das Verhalten in Schere – Stein – Papier und über die zugehörigen Spieler zur Verfügung stünden, dann könnte eine künstliche Intelligenz allerlei Abweichungen vom Rationalverhalten feststellen. Aus der Verhaltensforschung wissen wir zum Beispiel, dass Menschen Schwierigkeiten haben, unabhängig von vergangenem Verhalten Entscheidungen zu treffen. Wenn in den vergangenen Runden bereits Schere gewählt wurde, wird es etwa unwahrscheinlicher, dass auch in der nächsten Runde nochmals Schere gewählt wird. Dieses Verhaltensphänomen ist als „Gesetz der kleinen Zahl“ bekannt. Es ist irrational und kann daher durch eine kluge datengetriebene Prognose über die nächste Aktion des Gegenübers ausgebeutet werden.

Die Strategie eines menschlichen Spielers kann aber auch von soziodemographischen und anderen Charakteristiken abhängen, etwa wie viel Erfahrung der Spieler in diesem Spiel oder in ähnlichen Spielen besitzt, von der Tageszeit, ob Freunde zuschauen, wie viel Alkohol getrunken wurde, von dem Gemütszustand, vom Wetter und vielem mehr. Eine künstliche Intelligenz, die Zugriff zu einem reichen Datenschatz hat, kann solche Zusammenhänge identifizieren und dadurch Prognosen erstellen, die dabei helfen, Schere – Stein – Papier *systematisch* zu gewinnen. Wenn die Maschine

etwa prognostiziert, dass Stein am wahrscheinlichsten gewählt werden wird, wird sie mit Papier antworten.

Die perfekte Strategie weicht in diesem Fall von der rationalen Strategie (ein Drittel Wahrscheinlichkeit für jede Option) ab. Die Abweichung von der Rationalität des Menschen steckt die Maschine an, die aber ihrerseits nicht irrational, sondern superrational wird. Dadurch macht sich zwar auch die Maschine verwundbar. (Wenn nämlich der Mensch die Maschine durchschaut und erwartet, dass sie Papier spielt, dann kann er mit Schere den Gewinn erzwingen.) Doch spätestens seit Menschen Maschinen beim Schach hoffnungslos unterlegen sind, dürfte offenbar sein, dass Menschen an die Denktiefe einer mit vielen Daten gefütterten und mit enormer Rechenkapazität ausgestatteten intelligenten Maschine in der Regel nicht herankommen.

C. Mensch versus Maschine

Herbert A. Simon, Wirtschaftsnobelpreisträger und einer der Begründer der modernen Wissenschaft der Künstlichen Intelligenz, erklärte vor Jahrzehnten die Wirtschaftswissenschaft mit ihrem Fokus auf Rationalität zu einer Wissenschaft vom Künstlichen (“Economics is one of the sciences of the artificial”).³ Wenn künstliche Intelligenzen mit enormer Rechenkapazität aufeinandertreffen und miteinander interagieren, ist Rationalität ein nützliches Konzept, um Strategien und Ergebnisse zu prognostizieren. Es gibt auch eine Welt der Mathematiker und Spieltheoretiker, in denen das Rationalitätskonzept unmittelbar weiterhilft. In großen Auktionen der Telekommunikationsindustrie etwa, auf Finanzmärkten oder bei militärischen Konflikten, wenn Experten unablässig analysieren, wie die Wettbewerber ticken und welche Strategie wohl rational ist. Doch in einer Welt, in der Menschen aus Fleisch und Blut mitmischen, mit all ihren kognitiven und motivationalen Beschränkungen, führt die Rationalprognose für sich genommen selten zur besten Prognose. Datenbasierte Ansätze wie die experimentelle Forschung und, bei großen Datenmengen, die künstliche Intelligenz können dann die Prognosegüte verbessern helfen und so zu klügeren Entscheidungen führen.

Es ist vielleicht weniger überraschend, dass uns die künstliche Intelligenz durch neue Technologien und den Einsatz massiver Computerpower in Spielen wie Schach überlegen ist. Diese Intelligenz, richtig angewendet, eröffnet neue Möglichkeiten bei der Bewältigung unserer Herausforderungen. Doch

3 H. A. Simon, *Models of Bounded Rationality*, Cambridge, Mass., MIT Press 1976.

eine besondere Faszination für die künstliche Intelligenz entsteht, weil die menschliche Intelligenz so beschränkt ist. Intelligente Maschinen verstehen die Menschen mit ihren Irrationalitäten und Schwächen in mehr und mehr ökonomischen und sozialen Kontexten besser und können uns daher besser prognostizieren, als wir uns selbst verstehen und unsere eigenen Handlungen vorhersagen können. Dies führt dazu, dass die künstliche Intelligenz Menschen selbst in einfachsten Situationen (wie Schere – Stein – Papier) gut einschätzen kann und uns dadurch systematisch überlegen ist. Dazu kommt, dass Akteure, die gelernt haben, wie Menschen auf soziale oder ökonomische Impulse, auf Entscheidungsarchitekturen oder auf Informationsfeedback reagieren, auch Verhalten in eine gewünschte Richtung lenken können. Das muss nicht negativ sein. Viele Herausforderungen erfordern ein verändertes Verhalten – man denke etwa an den menschengemachten Klimawandel, Pandemien, Verkehrskollaps, Selbstkontrollprobleme und Finanzmarktexzesse. Doch auch die Sorge vor ökonomischem Missbrauch, Ausbeutung und vor einer Aushöhlung der Demokratie durch Eingriffe und Manipulationen ist berechtigt.

Eine breite Diskussion darüber, wie und unter welchen Umständen moderne Formen der Verhaltenslenkung eingeschränkt werden sollen, ist notwendig. Hier steht die Forschung erst am Anfang. Ein zu Regulierung und Gesetzgebung ergänzender Weg, solchen Herausforderungen zu begegnen, ist die Entwicklung von Regeln und Algorithmen, die wünschenswerte Ergebnisse aus sich heraus erreichen, selbst wenn sich alle (algorithmischen und menschlichen) Spieler bestmöglich im jeweiligen individuellen Eigeninteresse verhalten. Viele Beispiele im Forschungsfeld „Behavioral Market Design“ zeigen, dass dieser Weg vielversprechend ist. Auch hier spielen Spieltheorie und künstliche Intelligenz eine zentrale Rolle.

Auf dem Sprung zum autonomen Fahren

Martin Paul Waßmer

A. Einführung

Zu den zahlreichen Anwendungsfeldern der digitalen Transformation gehört der Verkehr. Auch wenn wir noch sehr weit von Programmen entfernt sind, die eine menschenähnliche künstliche Intelligenz aufweisen, sind bereits heute vielfach Assistenzsysteme im Einsatz. Dieser Trend wird sich verstärken, wodurch die Verkehrssicherheit erhöht, und die Zahl der Unfälle zurückgehen wird.¹ Denn menschliches Fehlverhalten ist der größte Risikofaktor im Straßenverkehr. Im Jahr 2020² gingen 88,5 % aller Unfälle mit Personenschaden auf das Fehlverhalten von Fahrzeugführern zurück. Technische Mängel und Wartungsmängel waren nur für 1,0 % der Unfälle ursächlich.

Allerdings ist die Automatisierung nicht ohne Risiken. In Deutschland führte im Jahr 2012 der Einsatz eines Spurhalteassistenten in Alzenau bei Aschaffenburg zum Tod einer Mutter und ihres siebenjährigen Sohnes.³ Nachdem der Fahrer eines Kfz einen Schlaganfall erlitten und das Bewusstsein verloren hatte, führte der Spurhalteassistent das Fahrzeug mit hoher Geschwindigkeit durch den Ort, wo es zur Kollision kam. Ohne den Assistenten wäre das Fahrzeug bereits am Ortseingang zum Stillstand gekommen. In den USA war im März 2018 ein autonom fahrender Testwagen von Uber in einen tödlichen Unfall verwickelt. Eine Fußgängerin, die ein Fahrrad schob, überquerte außerhalb des Zebrastreifens eine Straße, wurde erfasst und starb.⁴

1 BT-Drs. 19/27439, S. 15; G. M. Sander/J. Hollering, Strafrechtliche Verantwortlichkeit im Zusammenhang mit automatisiertem Fahren, *NStZ* 2017, 193.

2 *Statistisches Bundesamt*, Fachserie 8 Reihe 7: Verkehr, Verkehrsunfälle 2020, Juli 2021, S. 49.

3 Hierzu E. Hilgendorf, Automatisiertes Fahren und Strafrecht – der „Aschaffener Fall“, *DRiZ* 2018, 66, (68 f.).

4 A. Spies, USA: Autonom gesteuerter PKW verursacht schweren Unfall in Arizona, *ZD-Aktuell* 2018, 06029.

B. Entwicklungsstufen

Die internationale Norm SAE J 3016 definiert sechs Stufen.⁵ Während bei Stufe 0 (No Automation – keine Automatisierung) ein Mensch die Fahraufgabe vollständig übernimmt, helfen auf Stufe 1 (Driver Assistance – Fahrassistenz) einfache Assistenzsysteme (z.B. Abstandsregeltempomat) bei der Bedienung mit. Bei Stufe 2 (Partial Automation – Teilautomatisierung) werden diffizilere Funktionen übernommen (z.B. Spurhalten). Auf Stufe 3 (Conditional Automation – Bedingte Automatisierung) führt das Kfz selbstständig komplexe Funktionen aus (z.B. Spurwechsel), wobei der Fahrer das System nicht mehr ständig überwachen muss. Auf Stufe 4 (High Automation – Hochautomatisierung) übernimmt das System dauerhaft die Führung des Fahrzeugs in festgelegten Gebieten und unter bestimmten Bedingungen. Kann das System die Fahraufgabe nicht mehr bewältigen, wird der Fahrer aufgefordert, die Führung zu übernehmen. Reagiert er nicht, wird das Fahrzeug in einen risikominimalen Systemzustand versetzt. Schließlich gibt es bei Stufe 5 (Full Automation – Vollautomatisierung) nur noch Passagiere.

Eine abweichende Einteilung hatte das Bundesamt für Straßenwesen im Jahr 2012 vorgenommen, indem es – ohne die Stufen zu nummerieren und ohne das autonome Fahren zu erwähnen – die Fähigkeiten der SAE-Stufe 3 als „hochautomatisiert“ und die der SAE-Stufe 4 als „vollautomatisiert“ bezeichnete.⁶ Dieser Kategorisierung machte sich nachfolgend der vom Verkehrsministerium eingesetzte „Runde Tisch Automatisiertes Fahren“ zu eigen, und sie bildete sodann auch die Grundlage für die deutsche Gesetzgebung.⁷ In internationalen Zusammenhängen ist diese Einteilung jedoch unüblich – und sie hat auch in Deutschland zu Missverständnissen geführt.⁸

C. Rechtsrahmen

Hinsichtlich des Rechtsrahmens ist zu differenzieren. Unproblematisch sind die SAE-Stufen 1 und 2. Die Fahrassistenz und die Teilautomatisierung stehen mit dem deutschen Straßenverkehrsrecht und den internationalen

5 <https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>.

6 Siehe T. M. Gasser *et al.*, Rechtsfolgen zunehmender Fahrzeugautomatisierung, Berichte der Bundesanstalt für Straßenwesen, Heft F 83, Januar 2012, S. 9.

7 *Wissenschaftliche Dienste Deutscher Bundestag*, Autonomes und automatisiertes Fahren auf der Straße – rechtlicher Rahmen, WD7 – 3000 – 111/18, S. 4, Fn. 1.

8 Krit. H. Steege, Gesetzesentwurf zum autonomen Fahren (Level 4), SVR 2021, 128 (130).

Vorgaben des Wiener Übereinkommens über den Straßenverkehr⁹ in Einklang.¹⁰ Der Fahrer führt weiterhin das Kfz, da ihm die Assistenzsysteme nur Hilfestellungen geben. Dagegen wird in den SAE-Stufen 3 und 4 ein Grad an Automatisierung erreicht, bei dem das Kfz durch eine Fahrfunktion gesteuert wird und der Fahrer das System nicht mehr permanent überwachen muss.

I. Ahtes Gesetz zur Änderung des Straßenverkehrsgesetzes vom 16. Juni 2017

Der deutsche Gesetzgeber hatte dem „hochautomatisierten“ und „vollautomatisierten“ Fahren – d.h. dem automatisierten Fahren der SAE-Stufen 3 und 4 – bereits vor rund vier Jahren mit dem „Achten Gesetz zur Änderung des Straßenverkehrsgesetzes“ vom 16. Juni 2017 (BGBl. I, S. 1648) Rechnung getragen, das am 21. Juni 2017 in Kraft trat. Es handelte sich um den weltweit ersten legislativen Vorstoß dieser Art.

Danach ist der Betrieb eines Kfz mit hoch- oder vollautomatisierter Fahrfunktion zulässig, wenn die Funktion bestimmungsgemäß verwendet wird (§ 1a Abs. 1 StVG). Die Anforderungen sind hoch (§ 1a Abs. 2 StVG): Es muss eine Fahrzeugsteuerung vorhanden sein (Nr. 1), die in der Lage ist, den Verkehrsvorschriften zu entsprechen (Nr. 2), die jederzeit manuell übersteuerbar oder deaktivierbar ist (Nr. 3), die die Erforderlichkeit der eigenhändigen Fahrzeugsteuerung erkennen kann (Nr. 4), dies mit ausreichender Zeitreserve wahrnehmbar anzeigen kann (Nr. 5) und auf eine der Systembeschreibung zuwiderlaufende Verwendung hinweist (Nr. 6). Klargestellt wurde, dass auch derjenige, der eine hoch- oder vollautomatisierte Fahrfunktion aktiviert und verwendet, Fahrzeugführer ist, selbst wenn er im Rahmen der bestimmungsgemäßen Verwendung das Kfz nicht eigenhändig steuert (§ 1a Abs. 4 StVG). Weiter wurde klargestellt, dass sich der Fahrzeugführer vom Verkehrsgeschehen und der Fahrzeugsteuerung abwenden darf (§ 1b Abs. 1 Hs. 1 StVG). Dabei muss er jedoch derart wahrnehmungsbereit bleiben, dass er „jederzeit“ die Steuerung „unverzüglich“ wieder übernehmen kann, wenn ihn das System dazu auffordert oder er erkennt oder auf Grund offensichtlicher Umstände

9 Vienna Convention on Road Traffic vom 8.11.1968, United Nations, Treaty Series, Bd. 1042, S. 17.

10 U. Franke, Rechtsprobleme beim automatisierten Fahren – ein Überblick, DAR 2016, 61 (64); U. Lange, Automatisiertes und autonomes Fahren – eine verkehrs-, wirtschafts- und rechtspolitische Einordnung, NZV 2017, 345 (348).

erkennen muss, dass die Voraussetzungen für eine bestimmungsgemäße Verwendung nicht mehr vorliegen (§ 1b Abs. 1 Hs. 2, Abs. 2 Nr. 1 und 2 StVG).

Mit diesen Änderungen hatte der Gesetzgeber die Voraussetzungen festgelegt, unter denen der Einsatz der entsprechenden Fahrfunktionen im Straßenverkehr möglich ist. Allerdings war der Einsatz weiterhin nicht zulässig, da das technische Zulassungsrecht den Betrieb derart ausgerüsteter Fahrzeuge noch nicht gestattete.

II. Das „Gesetz zum autonomen Fahren“ vom 12. Juni 2021

Mit dem „Gesetz zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes – Gesetz zum autonomen Fahren“ vom 12. Juni 2021 (BGBl. I, S. 3108), das am 28. Juni 2021 in Kraft trat, wurde der nationale Rechtsrahmen erweitert. Hiermit wollte die Bundesregierung das ambitionierte Ziel erreichen, dass „Deutschland der erste Staat wird, der Fahrzeuge ohne Fahrer aus der Forschung in den Alltag holt“.¹¹

Vorausgegangen waren Änderungen und Ergänzungen der UN/ECE-Regelungen, sowie Änderungen des Wiener Übereinkommens über den Straßenverkehr. So ermöglicht nunmehr die UN/ECE-Regelung Nr. 79 (Steering Equipment) den Einsatz automatischer Lenkfunktionen und die UN/ECE-Regelung Nr. 157 (Automated Lane Keeping Systems) die Genehmigung von Autobahnstapiloten, die bis 60 km/h die Fahrzeugführung übernehmen.¹² Nach dem neuen Art. 34bis des Wiener Übereinkommens gilt das Erfordernis, dass jedes Fahrzeug einen Führer haben muss, für automatische Fahrsysteme als erfüllt, wenn diese den nationalen Vorschriften für den Betrieb entsprechen. Damit wurde klargestellt, dass das Übereinkommen auch das autonome Fahren gestattet.¹³

Das Gesetz zum autonomen Fahren richtet sich in Anbetracht der hohen Anforderungen weniger an den Individualverkehr, sondern in erster Linie an

11 <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/gesetz-zum-autonomen-fahren.html>.

12 L. S. Lutz, Neue Vorschriften für das automatisierte und autonome Fahren – ein Überblick, DAR 2021, 182 (182 f.).

13 S. Gatzke, Gesetz zum autonomen Fahren – Ist die externe Überwachung autonomer Fahrsysteme mit dem Wiener Übereinkommen über den Straßenverkehr vereinbar?, NZV 2021, 402 (406 f.).

Personen- und Gütertransportanbieter.¹⁴ Eingesetzt werden sollen Kfz mit autonomer Fahrfunktion im Straßenverkehr und auf Werksgeländen im Shuttle-Verkehr auf festgelegten Routen, im Hub2Hub-Verkehr zwischen Verteilzentren, zur Beförderung von Personen und/oder Gütern auf der ersten oder letzten Meile sowie beim vollautomatisierten Parkservice. Die neuen §§ 1d ff. StVG regeln die technischen Anforderungen, das Verfahren für die Erteilung einer Betriebserlaubnis, die Pflichten der am Betrieb beteiligten Personen, die Datenverarbeitung sowie die (nachträgliche) Aktivierung von automatisierten und autonomen Fahrfunktionen bereits typgenehmigter Kraftfahrzeuge. Darüber hinaus wurden Vorschriften geschaffen, welche die Erprobung ermöglichen.

Gestattet wird der Einsatz eines Kfz mit autonomer Fahrfunktion, wenn es die Fahraufgabe ohne eine fahrzeugführende Person selbstständig in einem festgelegten Betriebsbereich erfüllen kann und über eine entsprechende technische Ausrüstung verfügt (§ 1d Abs. 1 Nr. 1 und 2 StVG). Gefordert wird eine sog. Technische Aufsicht, die aus einer natürlichen Person, also einem Menschen besteht (§ 1d Abs. 3 StVG). Damit hat sich der Gesetzgeber dagegen entschieden, die Aufsicht einer Künstlichen Intelligenz anzuvertrauen. Hierfür dürfte nicht nur maßgebend gewesen sein, dass ein Mensch die Kontrolle behalten soll, sondern auch, dass im deutschen Recht bislang nur Menschen Träger von Rechten und Pflichten sind und damit haften können.¹⁵ Die Technische Aufsicht hat keinen Zugriff auf die Fahrzeugsteuerung, muss das Fahrzeug aber von außen jederzeit deaktivieren und erforderlichenfalls alternative Fahrmanöver freischalten können, die das System dann eigenständig durchführt (§ 1f Abs. 2 StVG). Die Anforderungen, welche die natürliche Person erfüllen muss,¹⁶ werden durch eine Rechtsverordnung¹⁷ geregelt (§ 1j Abs. 1 Nr. 4 lit. a StVG). Ein weitergehender Einsatz im Straßenverkehr ist mit einer Erprobungsgenehmigung möglich (§ 1i StVG).

Angesichts der Restriktionen wird deutlich, dass es nicht um das autonome Fahren der SAE-Stufe 5 geht, sondern nur um die SAE-Stufe 4¹⁸ und damit

14 S. Kleemann/C. Arzt, Das Gesetz zum „autonomen“ Fahren in Deutschland, RAW 2021, 99 (104).

15 E. Hilgendorf, Straßenverkehrsrecht der Zukunft, JZ 2021, 444 (446).

16 T. Haupt, Auf dem Weg zum autonomen Fahren, Der Entwurf eines Gesetzes zum autonomen Fahren, NZV 2021, 172 (175).

17 Verordnung zur Regelung des Betriebs von Kraftfahrzeugen mit automatisierter und autonomer Fahrfunktion und zur Änderung straßenverkehrsrechtlicher Vorschriften vom 24. Juni 2022, BGBl. I S. 986.

18 So auch BT-Drs. 19/27439, S. 16.

um ein „beschränkt“ autonomes Fahren.¹⁹ Konsequenz wäre es daher gewesen, vom „Gesetz zum hochautomatisierten Fahren“ zu sprechen und aus den §§ 1a, 1b StVG den Begriff „vollautomatisiert“ zu streichen.²⁰ Andererseits ist aber in den entsprechend ausgestatteten Kfz kein Fahrer mehr vorhanden. Zudem ist mit einer Erprobungsgenehmigung ein über festgelegte räumliche Bereiche hinausgehender Einsatz im Straßenverkehr „zur Erprobung von Entwicklungsstufen für die Entwicklung automatisierter oder autonomer Fahrfunktionen“ (§ 1i Abs. 1 StVG) zulässig. Daher ist die Bezeichnung als „Gesetz zum autonomen Fahren“ noch vertretbar. Sichtbar wird allerdings, dass es sich um eine „Übergangslösung“²¹ handelt und die gesetzliche Terminologie der Anpassung an die internationale Norm SAE J 3016 bedarf.

D. Zivilrechtliche Haftung

Am zivilrechtlichen Haftungsregime hat sich durch die Neuregelungen nichts grundlegend geändert. Im deutschen Recht existiert mit der verschuldensunabhängigen Halterhaftung (§ 7 StVG) eine Gefährdungshaftung, die allein an die Betriebsgefahr anknüpft. Der Halter hat hierbei auch für Fehler der Fahrfunktion einzustehen.²² Die Haftungshöchstgrenzen bei Verwendung einer Fahrfunktion wurden verdoppelt (§ 12 Abs. 1 StVG). Zudem ist der Halter eines Kfz mit autonomer Fahrfunktion verpflichtet, eine Haftpflichtversicherung „auch für eine Person der Technischen Aufsicht abzuschließen und aufrechtzuerhalten“ (§ 1 S. 2 PflVG). Weiter besteht eine verschuldensabhängige Fahrerhaftung (§ 18 StVG). Der Fahrer kann die Verschuldensvermutung entkräften, indem er einen nicht erkennbaren unfallkausalen Fehler des Systems nachweist.²³

Die Technische Aufsicht haftet dagegen nach § 823 Abs. 1 BGB verschuldensabhängig. Der Empfehlung des Bundesrats, eine Gefährdungshaftung oder zumindest eine Haftung für vermutetes Verschulden zu normieren,²⁴

19 Hilgendorf, Straßenverkehrsrecht (Fn. 15), 445.

20 Siehe Kleemann/Arzt, Gesetz (Fn. 14), 100 f.; Steege, Gesetzentwurf (Fn. 8), 130.

21 BT-Drs. 19/25626, S. 6; M. Wagner, Gesetz zum autonomen Fahren – Streitpunkte im Gesetzgebungsverfahren, SVR 2021, 287 (292).

22 F. P. Schuster, Strafrechtliche Verantwortlichkeit der Hersteller beim automatisierten Fahren, DAR 2019, 6.

23 R. Greger, Haftungsfragen beim automatisierten Fahren, NZV 2018, 1 (2).

24 BT-Drs. 19/28178, S. 11.

folgte die Bundesregierung nicht, da sie die Haftung im Rahmen des allgemeinen Deliktsrechts als ausreichend betrachtete.²⁵

Schließlich haftet der Hersteller im Wege der Produkthaftung verschuldensunabhängig nach dem ProdHaftG für Folgeschäden aus dem Gebrauch. Der Anspruchssteller muss nur nachweisen, dass ein Fehler des Produkts – d.h. vor allem eine fehlerhafte Programmierung der Fahrfunktion, aber auch ein Konstruktions-, Fabrikations- oder Instruktionsfehler²⁶ – vorlag und daraus ein kausaler Schaden entstanden ist (§ 1 Abs. 4 ProdHaftG). Bei der Beweisführung kommt ihm ein Anscheinsbeweis zugute, sofern ein typischer Geschehensablauf vorliegt. Darüber hinaus haftet der Hersteller ergänzend im Rahmen der Produzentenhaftung gem. §§ 823 ff. BGB verschuldensabhängig für die Eröffnung und Beherrschung einer Gefahrenquelle. Insoweit besteht eine Beweislastumkehr, d.h. der Geschädigte muss lediglich beweisen, dass das Produkt fehlerhaft war, als es den Geschäftskreis des Herstellers verlassen hat, und dass zwischen Fehlerhaftigkeit und Schaden ein Kausalzusammenhang besteht.²⁷

Perspektivisch ist zu erwarten, dass sich mit zunehmender Automatisierung die zivilrechtliche Haftung vom Halter und Fahrer auf den Hersteller verlagert.²⁸ Gleichzeitig dürfte die Anzahl der Unfälle immer weiter zurückgehen, womit neben den Herstellern auch die Fahrer und Halter entlastet werden, da sie geringere Versicherungsprämien aufbringen müssen. Damit profitieren im Ergebnis alle Beteiligten.

E. Strafrechtliche Verantwortlichkeit

Die Neuregelungen haben die Grundsätze der strafrechtlichen Verantwortlichkeit ebenfalls nicht verändert. Mit Blick auf das verfassungsrechtlich verankerte Schuldprinzip²⁹ ist eine verschuldensunabhängige Gefährdungshaftung ausgeschlossen. Strafrechtliche Verantwortung erfordert stets den Nachweis individuellen Verschuldens.

25 Schuster, Verantwortlichkeit (Fn. 22), 6.

26 V. M. Jänich/P. T. Schrader/V. Reck, Rechtsprobleme des autonomen Fahrens, NZV 2015, 313 (316 f.).

27 Jänich/Schrader/Reck, Rechtsprobleme (Fn. 26), 318.

28 E. Hilgendorf, Automatisiertes Fahren und Recht – ein Überblick, JA 2018, 801 (803).

29 S. nur BVerfGE 20, 323 (331).

I. Verantwortlichkeit für Körperverletzungs- und Tötungsdelikte

Im Vordergrund steht beim Fahren mit aktivierter Fahrfunktion die Verantwortlichkeit aus Fahrlässigkeitsdelikten, vor allem wegen fahrlässiger Körperverletzung (§ 229 StGB) und fahrlässiger Tötung (§ 222 StGB).³⁰ Geboten ist eine Doppelprüfung. Es muss nicht nur eine objektive Sorgfaltspflichtverletzung vorliegen und der Eintritt des Erfolgs objektiv vorhersehbar gewesen sein, sondern es ist darüber hinaus festzustellen, dass der Täter Sorgfaltspflichtverletzung und Erfolgseintritt nach seinen individuellen Fähigkeiten und dem Maß seines individuellen Könnens hätte vermeiden können. Der Kausalitätsnachweis ist beim Einsatz von Kfz mit hoch- oder vollautomatisierter Fahrfunktion wegen § 63a StVG, der eine Datenspeicherung („Blackbox“) vorschreibt, weniger problematisch als sonst.³¹ In Bezug auf Kfz mit autonomer Fahrfunktion wurde diese Regelung nicht erweitert.³² Die Fahrzeuge verfügen aber ohnehin über einen Datenspeicher (§ 1g StVG).

Beim automatisierten Fahren ist es denkbar, dass der Fahrer ein System aktiviert und nutzt, bei dem er nicht erkannt hat (aber hätte erkennen können), dass es nicht ordnungsgemäß funktioniert, oder dass er es unsachgemäß bedient.³³ Hat der Fahrer dagegen erkannt, dass das System eine Fehlfunktion aufweist und fährt er dennoch, liegt die Annahme von Eventualvorsatz nahe, womit sogar eine Verantwortlichkeit aus einem Vorsatzdelikt – wegen Sachbeschädigung (§ 303 StGB), Körperverletzung (§§ 223 ff. StGB) oder Totschlag (§ 212 StGB) – in Frage kommt.

Der Halter eines Kfz mit autonomer Fahrfunktion ist zur Erhaltung der Verkehrssicherheit verpflichtet und hat u.a. die regelmäßige Wartung der erforderlichen Systeme sicherzustellen (§ 1f Abs. 1 StVG). Entsprechendes gilt für Fahrzeuge mit hoch- oder vollautomatisierter Fahrfunktion.³⁴ Daher kann eine Strafbarkeit wegen eines Wartungs- oder Funktionsmangels bestehen. In Betracht kommen aber im Hinblick auf die Aktivierung der Fahrfunktion auch Instruktionmängel. Zudem ist es denkbar, dass die Technische Aufsicht fehlerhaft ausgewählt und kontrolliert wurde.³⁵

30 Sander/Hollering, Verantwortlichkeit (Fn. 1), 195 f.

31 Schuster, Verantwortlichkeit (Fn. 22), 7.

32 Krit. Steege, Gesetzentwurf (Fn. 8), 135.

33 Vgl. N. Nestler, Überlegungen zum Umgang mit Kraftfahrzeugen mit automatisierter oder autonomer Fahrfunktion im (Straßenverkehrs-)Strafrecht, Jura 2021, 1183 (1192).

34 S. nur C. Staub, Strafrechtliche Fragen zum Automatisierten Fahren, NZV 2019, 392 (396).

35 Hilgendorf, Straßenverkehrsrecht (Fn. 15), 453.

Die natürliche Person, die als Technische Aufsicht fungiert, kann sich ebenfalls strafbar machen. Sie ist zwar nicht zur ständigen Überwachung verpflichtet, wohl aber zur Evidenzkontrolle. Sie muss die Signale der Fahrfunktion bewerten und entsprechend reagieren können. Schlägt die Fahrfunktion in einer die Fahraufgabe beeinträchtigenden Lage, die nicht eigenständig bewältigt werden kann, ein alternatives Fahrmanöver vor, hat die Technische Aufsicht das Fahrmanöver zu bewerten und freizugeben.³⁶ Ferner darf sie Vorschläge zu Fahrmanövern, die abstrakt geltende Vorschriften verletzen würden, als zulässig freigeben (z.B. die Umfahrung eines verunfallten Fahrzeugs; das Weiterfahren bei einer defekten Ampelanlage mit „Dauerrot“).³⁷ Schließlich hat die Technische Aufsicht im Fall der Deaktivierung, wenn das Fahrzeug in einen risikominimalen Zustand versetzt wird, sicherzustellen, dass das Fahrzeug kein Verkehrshindernis darstellt. Auch hierbei können den eingesetzten Menschen Fehler unterlaufen.

Eine strafrechtliche Verantwortlichkeit der Passagiere ist dagegen beim autonomen Fahren ausgeschlossen, da sie das Kfz nicht führen.³⁸ Eine Verantwortlichkeit kommt erst in Betracht, wenn sie das Kfz ausnahmsweise tatsächlich beherrschen und steuern können.

Der Schwerpunkt der strafrechtlichen Verantwortlichkeit liegt beim automatisierten und autonomen Fahren beim Hersteller, sofern ihm ein Konstruktions-, Fabrikations- oder Instruktionsfehler oder die Verletzung von Produktbeobachtungs- und Rückruffpflichten vorgeworfen werden kann.³⁹ Maßgebend ist dabei, ob einer für den Hersteller tätigen natürlichen Person, insbesondere dem Leiter der Entwicklungsabteilung, aber auch einem Konstrukteur, Programmierer oder Techniker, ein konkreter Vorwurf zu machen ist. Lassen sich die innerbetrieblichen Vorgänge nicht mehr rekonstruieren, ist eine Aufsichtspflichtverletzung nach § 130 OWiG naheliegend.⁴⁰ Hat insoweit eine Leitungsperson ihre Pflichten verletzt, kann eine Verbandsgeldbuße (§ 30 OWiG) festgesetzt werden. Die aktivierten Fahrfunktionen müssen jedoch nicht absolut, sondern nur relativ sicher sein. Unter dem Aspekt des erlaubten Risikos sind höhere Risiken hinnehmbar, sofern sie so weit wie

36 BT-Drs. 27439, S. 25.

37 BT-Drs. 27439, S. 25; *Hilgendorf*, Straßenverkehrsrecht (Fn. 15), 448.

38 *Schuster*, Verantwortlichkeit (Fn. 22), 6.

39 Siehe nur *Sander/Hollering*, Verantwortlichkeit (Fn. 1), 197 ff.

40 *Schuster*, Verantwortlichkeit (Fn. 22), 9.

möglich minimiert werden.⁴¹ Entscheidend ist, ob die Risikobilanz insgesamt positiv ist, da Fehler angesichts der Komplexität des Straßenverkehrs nicht auszuschließen sind.⁴²

Schließlich ist es denkbar, dass ein Kfz-Betrieb die aktivierte Fahrfunktion unsachgemäß eingebaut, nachgerüstet, angepasst, repariert oder gewartet hat.

II. Verantwortlichkeit für Straßenverkehrsdelikte

Darüber hinaus kann eine Verantwortlichkeit für Straßenverkehrsdelikte bestehen. Teilweise wird angenommen, der Fahrer führe beim automatisierten Fahren das Fahrzeug nicht, sondern habe die Fahrzeugführung mit der Inbetriebnahme abgegeben.⁴³ Da es sich bei § 315c StGB (Gefährdung des Straßenverkehrs) um ein eigenhändiges Delikt handelt, müsste demzufolge eine Verantwortlichkeit ausscheiden. Andererseits ist aber nach § 1a Abs. 4 StVG auch derjenige Fahrzeugführer, der die Fahrfunktion aktiviert und zur Fahrzeugsteuerung verwendet, selbst wenn er das Kfz nicht eigenhändig steuert. An diese außerstrafrechtliche Klarstellung ist grds. anzuknüpfen. Da für das Strafrecht aber die tatsächliche Beherrschung maßgebend ist, liegt ein „Führen“ nur dann vor, wenn der Fahrer die Fahrzeugsteuerung tatsächlich überwacht und jederzeit übernehmen könnte bzw. sie wegen der durch das System signalisierten Erforderlichkeit der Übernahme übernehmen müsste. Zudem liegt ein Führen auch dann vor, wenn der Fahrer erkennt, dass das System nicht ordnungsgemäß funktioniert. In diesen Fällen kommt eine Strafbarkeit in Betracht, wenn der Fahrer gerade wegen oder trotz des Einsatzes des Systems grob verkehrswidrig und rücksichtslos im Straßenverkehr eine der sog. „sieben Todsünden“ vorsätzlich begeht (§ 315c Abs. 1 Nr. 2 lit. a bis g StGB). Strafbar ist auch die fahrlässige Begehung (§ 315c Abs. 3 StGB), insbesondere wenn sich der Fahrer aus Gleichgültigkeit keine Gedanken macht.⁴⁴ Darüber hinaus kann sich der Fahrer aus § 315b StGB (Gefährliche Eingriffe in den

41 E. Hilgendorf, Dilemma-Probleme beim automatisierten Fahren, ZStW 130 (2018), 674 (700); Sander/Hollering, Verantwortlichkeit (Fn. 1), 197; Staub, Strafrechtliche Fragen (Fn. 34), 397.

42 Bundesministerium für Verkehr und digitale Infrastruktur (Hrsg.), Bericht der Ethik-Kommission „Automatisiertes und Vernetztes Fahren“, Juni 2017, S. 10; T. Schulz, Sicherheit im Straßenverkehr und autonomes Fahren, NZV 2017, 548 (551).

43 Nestler, Überlegungen (Fn. 33), 1190.

44 T. Fischer, Strafgesetzbuch mit Nebengesetzen, 69. Aufl., München 2022, § 315c Rn. 14.

Straßenverkehr) strafbar machen, wenn er einen sog. „verkehrsfeindlichen Inneneingriff“ begeht, d.h. das Kfz unter Einsatz der Fahrfunktion zur Schädigung anderer Verkehrsteilnehmer oder fremder Sachen einsetzt.⁴⁵

Eine Strafbarkeit der Passagiere aus einem Straßenverkehrsdelikt scheidet dagegen beim autonomen Fahren aus, da das Kfz allein durch das System geführt wird.⁴⁶

Die natürliche Person, die als Technische Aufsicht fungiert, kann nicht aus § 315c StGB strafbar sein, da sie das Kfz nicht eigenhändig führt, sondern es lediglich deaktivieren und Fahrmanöver freigeben kann, welche die Fahrfunktion dann eigenständig durchführt.⁴⁷ In Betracht kommt aber im Fall der Deaktivierung eine Strafbarkeit aus § 315b Abs. 1 Nr. 2 StGB, falls die Technische Aufsicht vorsätzlich oder fahrlässig (§ 315b Abs. 4, 5 StGB) nicht sicherstellt, dass das Kfz kein Hindernis darstellt. Denkbar ist, dass das Kfz so abgestellt wird, dass die Sicherheit des Straßenverkehrs stark beeinträchtigt ist. In Anbetracht dessen, dass die Technische Aufsicht nur die Deaktivierung einleitet und das System diese dann in Übereinstimmung mit den Verkehrsregeln eigenständig ausführt, wird jedoch eine Strafbarkeit durchweg auscheiden.

Eine Strafbarkeit der natürlichen Personen, die für den Hersteller bzw. einen Kfz-Betrieb handeln, aus § 315c StGB scheidet ebenfalls aus, da auch sie das Fahrzeug nicht eigenhändig führen. In Betracht kommt aber bei einer fehlerhaften Programmierung oder beim Einbau fehlerhafter Sensoren eine Strafbarkeit aus § 315b StGB.⁴⁸ Voraussetzung ist, dass das Fehlverhalten als „ähnlicher, ebenso gefährlicher Eingriff“ (§ 315b Abs. 1 Nr. 3 StGB) bewertet werden kann und dadurch im Einzelfall Leib oder Leben eines anderen Menschen (des Fahrers, der Insassen oder anderer Verkehrsteilnehmer) oder fremde Sachen „von bedeutendem Wert“ (mindestens 750 EUR)⁴⁹ konkret gefährdet werden. Vom BayObLG war allerdings im Jahr 1973 die pflichtwidrige Schlechterfüllung eines Reparaturauftrags nicht als „Eingriff“ in den Straßenverkehr bewertet worden, da dies voraussetzen soll, dass das Fahrzeug absichtlich beschädigt oder in schadhaftem Zustand belassen wird.⁵⁰ Hier-

45 BGHSt 48, 233 (237); *F. Zieschang* in: U. Kindhäuser/U. Neumann /H.-U. Paeffgen (Hrsg.), NomosKommentar Strafgesetzbuch, 5. Aufl., Baden-Baden 2017, § 315b Rn. 11 ff. m.w.N.

46 *Nestler*, Überlegungen (Fn. 33), 1188.

47 *Nestler*, Überlegungen (Fn. 33), 1189.

48 *Schuster*, Verantwortlichkeit (Fn. 22), 7.

49 BGH NSTZ 2011, 215.

50 BayObLGSt 1973, 198.

gegen spricht jedoch, dass auch fahrlässiges Handeln unter Strafe steht (§ 315b Abs. 5 StGB). Zudem ist bei Außeneingriffen die Feststellung einer „Verkehrsfreundlichkeit“ bzw. Sabotageabsicht nicht erforderlich.⁵¹ Schließlich hatte der BGH bereits im Oktober 1971⁵² die fehlerhafte Ausbesserung einer Schienenstrecke als gefährlichen Eingriff in den Bahnverkehr (§ 315 StGB) bewertet. Konsequenz ist, dass der Strafvorschrift des § 315b StGB, die derzeit eine untergeordnete Bedeutung hat, beim autonomen Fahren eine Schlüsselrolle zukommen könnte.⁵³

III. Dilemma-Situationen

Große öffentliche Aufmerksamkeit haben Dilemma-Situationen⁵⁴ erfahren, bei denen sich sowohl auf dem vorgesehenen Fahrweg als auch auf den möglichen Ausweichstrecken Menschen befinden. Diese Situationen dürften zwar künftig noch seltener als heute vorkommen, da die Vernetzung von Fahrzeugen und der Einsatz künstlicher Intelligenz eine vorausschauende Fahrweise und beim Auftreten akuter Gefahrensituationen Reaktionen im Bereich von Millisekunden gestatten, sie sind aber nicht auszuschließen. Es handelt sich gewissermaßen um die „dunkle Seite“ des autonomen Fahrens,⁵⁵ die eine besondere Faszination ausübt, weil Systeme und damit Algorithmen über Leben und Tod entscheiden können. Die Frage ist, wie die Programmierung vorzunehmen ist.

Umstritten ist bereits, ob es um eine Strafbarkeit aus einem Vorsatz- oder Fahrlässigkeitsdelikt geht. Nach h.M.⁵⁶ ist eine Strafbarkeit wegen Totschlags (§ 212 StGB) denkbar, da die Fahrfunktion entsprechend ihrer Programmierung reagiert. Soweit diese den Beteiligten – Herstellern, Haltern, Fahrern bzw. Passagieren – bekannt sei, könnten sie an der Tat beteiligt sein. Der Tod von Menschen werde beim Eintritt einer Dilemma-Situation nicht nur ernst-

51 *Fischer* (Fn. 44), § 315b Rn. 6.

52 BGHSt 24, 231 (233).

53 *Schuster*, Verantwortlichkeit (Fn. 22), 7.

54 Siehe nur *S. Beck* in: B. H. Oppermann/J. Stender-Vorwachs (Hrsg.), *Autonomes Fahren*, 2. Aufl., München 2020, Kap. 3.7 Rn. 46 ff.

55 *A. Hevelke/J. Nida-Rümelin*, Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle, *JWE* 19 (2014), 5 (6).

56 *A. Engländer*, Das selbstfahrende Kraftfahrzeug und die Bewältigung dilemmatischer Situationen, *ZIS* 2016, 608 (611 ff.); *Sander/Hollering*, Verantwortlichkeit (Fn. 1), 202; *Schuster*, Verantwortlichkeit (Fn. 22), 7, 10; *P. Weber*, Dilemmasituationen beim autonomen Fahren, *NZV* 2016, 249 (251 ff.).

haft für möglich gehalten, sondern auch billigend in Kauf genommen. Hierbei seien diejenigen, welche die Fahrfunktion im Straßenverkehr aktivieren und nutzen, infolge ihrer Tatherrschaft als Täter anzusehen. Dies sei beim automatisierten Fahren der Fahrer, beim autonomen Fahren der Halter. Die übrigen Beteiligten könnten nur Gehilfen sein. Nach der Gegenauffassung⁵⁷ soll allenfalls eine Strafbarkeit aus fahrlässiger Tötung (§ 222 StGB) in Betracht kommen, da zwar stets mit einem Unfall gerechnet werden müsse, die Beteiligten aber annehmen dürften, es werde „alles gut gehen“. Dies ist allerdings eine Illusion, da die Dilemma-Situationen sich gerade dadurch auszeichnen, dass Menschen zwangsläufig zu Schaden kommen.

Mit dem Gesetz zum autonomen Fahren hat der Gesetzgeber erstmals einen Rahmen für Dilemma-Situationen vorgegeben. Nach § 1e Abs. 2 Nr. 2 StVG müssen autonome Fahrzeuge über ein System der Unfallvermeidung verfügen, „das a) auf Schadensvermeidung und Schadensreduzierung ausgelegt ist, b) bei einer unvermeidbaren alternativen Schädigung unterschiedlicher Rechtsgüter die Bedeutung der Rechtsgüter berücksichtigt, wobei der Schutz menschlichen Lebens die höchste Priorität besitzt, und c) für den Fall einer unvermeidbaren alternativen Gefährdung von Menschenleben keine weitere Gewichtung anhand persönlicher Merkmale vorsieht“. Der Gesetzgeber hat sich hierbei an den Empfehlungen der Ethik-Kommission „Automatisiertes und Vernetztes Fahren“ orientiert, die der Bundesminister für Verkehr und digitale Infrastruktur im Jahr 2016 eingesetzt hatte.⁵⁸ Jedoch sind die gesetzlichen Vorgaben bei näherer Betrachtung unvollständig, was dem Umstand geschuldet sein dürfte, dass bei vielen Konstellationen das Meinungsspektrum weit auseinandergeht.

1. Beschädigung oder Zerstörung von fremden Sachen

Nach den gesetzlichen Vorgaben hat der Schutz menschlichen Lebens höchste Priorität (§ 1e Abs. 2 Nr. 2 lit. b StVG). Daher ist die Beschädigung oder Zerstörung von fremden Sachen, insbesondere eines geparkten Autos oder auch eines Tieres, vom Eigentümer hinzunehmen, wenn damit die Verletzung oder gar Tötung von Menschen vermieden wird.⁵⁹ Im Strafrecht hat das

57 E. Hilgendorf, *Autonomes Fahren im Dilemma, Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätigen Kollisionsvermeidungssystemen*, in: E. Hilgendorf (Hrsg.), *Autonome Systeme und neue Mobilität*, Baden-Baden 2017, S. 143 (164 ff.).

58 *Bundesministerium für Verkehr und digitale Infrastruktur*, Bericht (Fn. 42), S. 7.

59 Siehe nur *Engländer*, *Das selbstfahrende Kraftfahrzeug* (Fn. 56), 615.

zugrunde liegende Prinzip des überwiegenden Interesses in § 34 StGB Ausdruck gefunden. Danach ist eine Tat, die zur Abwendung einer gegenwärtigen, nicht anders abwendbaren Gefahr begangen wird, nicht rechtswidrig, „wenn das geschützte das beeinträchtigte Interesse wesentlich überwiegt“. Dementsprechend hatte auch die Ethik-Kommission empfohlen, Tier- oder Sachschäden in Kauf zu nehmen, wenn dadurch Personenschäden vermeidbar sind.⁶⁰ Soweit teilweise angemerkt wird, die Berufung auf § 34 StGB setze voraus, dass die Gefahrenlage vom Täter „wirklich“ wahrgenommen werde, was im Vorfeld bei einem Programmierer nicht der Fall sei,⁶¹ überzeugt dies nicht. Denn Täter kann nur der Fahrer bzw. Halter sein, der sich durch die Aktivierung und Nutzung der Fahrfunktion die Programmierung zu eigen macht. Auch am Rettungswillen fehlt es nicht, da die Programmierung gerade vorgenommen wurde, um den Tod von Menschen zu vermeiden.

2. Tötung von Menschen außerhalb des Fahrzeugs

a) Fälle alternativer Gefährdung

Hat das Ausweichmanöver zur Folge, dass zwar ein Mensch (etwa ein Kind) gerettet, aber hierfür alternativ in das Leben eines anderen Menschen auf dem Ausweichweg eingegriffen würde, scheidet wegen des Grundsatzes des absoluten Lebensschutzes eine Rechtfertigung nach § 34 StGB aus.⁶² Das Leben eines Menschen steht in der Werteordnung an oberster Stelle, eine Relativierung oder Verrechnung (etwa nach Alter, Geschlecht, körperlicher bzw. geistiger Konstitution oder Staatsangehörigkeit) ist unzulässig.⁶³

Klargestellt hat der Gesetzgeber, dass in den Fällen der alternativen Gefährdung eine qualitative Gewichtung anhand persönlicher Merkmale unzulässig ist (§ 1e Abs. 2 Nr. 2 lit. c StGB). Dies entspricht der Empfehlung der Ethik-Kommission.⁶⁴ Das Überfahren eines Menschen, der sich im Fahrweg

60 Bundesministerium für Verkehr und digitale Infrastruktur, Bericht (Fn. 42), S. 11, Regel Nr. 7.

61 W. Mitsch, Die Probleme der Kollisionsfälle beim autonomen Fahren, KriPoZ 2018, 70 (71 f.).

62 Siehe nur T. Weigend, Notstandsrecht für selbstfahrende Autos?, ZIS 2017, 599 (600).

63 Hilgendorf, Dilemma-Probleme (Fn. 41), 695; abweichend T. Hörnle/W. Wohlers, The Trolley Problem Reloaded, GA 2018, 12 (28), die das Lebensalter als zulässiges Kriterium bewerten.

64 Bundesministerium für Verkehr und digitale Infrastruktur, Bericht (Fn. 42), S. 11, Regel Nr. 9.

befindet, ist daher im Ergebnis straflos. Ein Ausweichen und Überfahren eines anderen Menschen würde hingegen eine Strafbarkeit wegen eines Tötungsdelikts nach sich ziehen.⁶⁵ Eine entsprechende Programmierung ist unzulässig.

Nicht ausdrücklich geregelt hat der Gesetzgeber die Fälle des quantitativen Überwiegens, in denen durch ein Ausweichen mehr Menschen gerettet werden könnten, als geopfert werden müssten. Damit sind die umstrittensten Konstellationen ungelöst geblieben.⁶⁶ Ein bekanntes Lehrbuch-Beispiel bildet der „Weichensteller-Fall“ aus dem Jahr 1951,⁶⁷ in dem durch das Umleiten eines Zuges mehr Menschen gerettet als Gleisarbeiter getötet werden. Im angloamerikanischen Raum wird die Problematik anhand einer Straßenbahn thematisiert („Trolley-Problem“).⁶⁸ An der Rechtswidrigkeit des Ausweichens ändert es jedoch nichts, wenn quantitativ gesehen mehr Menschen gerettet als geopfert werden könnten. Ein Ausweichen wäre daher strafbar. Dieses Ergebnis lässt sich damit begründen, dass gesicherte Rechtspositionen (ungefährdet zu sein) einen höheren Wert als bloße Expektanzen (gerettet zu werden) haben, mithin die Unterlassungspflicht und damit das Verletzungsverbot („töte keinen Menschen“) gewichtiger ist als die Handlungs- und damit die Hilfespflicht („rette mehrere Menschen“).⁶⁹

Dieses Ergebnis wird vielfach als unbefriedigend empfunden. Zur Abhilfe werden mehrere Wege beschritten. Erstens wird angenommen, Unterlassungspflichten hätten keinen Vorrang gegenüber Handlungspflichten, womit ein Wahlrecht bestünde.⁷⁰ Dieser Lösung steht aber entgegen, dass durch den Lenkimpuls ein bislang nicht gefährdeter Mensch sterben müsste. Das Recht kann aber von keinem Menschen verlangen, sich für andere aufzuopfern.⁷¹ Im Übrigen ist eine Verrechnung von Menschenleben ausgeschlossen.⁷²

Zweitens wird angenommen, dass zwei Unterlassungspflichten kollidieren.⁷³ Der Hersteller habe bei der Programmierung weit im Vorfeld nach

65 A.A. (übergesetzlicher entschuldigender Notstand in Anlehnung an §§ 33, 35 StGB) Hörnle/Wohlers, Trolley Problem (Fn. 63), 15 f.

66 Krit. Steege, Gesetzentwurf (Fn. 8), 131.

67 H. Welzel, Zum Notstandsproblem, ZStW 63 (1951), 47 (51).

68 Hörnle/Wohlers, Trolley Problem (Fn. 63), 12 m.w.N.

69 P. F. Schuster, Notstandsalgorithmen beim autonomen Fahrzeug, RAW 2017, 13 (15 f.).

70 Hierzu T. Zimmermann, Rettungstötungen, Baden-Baden 2009, S. 123 f.

71 U. Neumann in: U. Kindhäuser/U. Neumann /H.-U. Paeffgen (Hrsg.), NomosKommentar Strafgesetzbuch, 5. Aufl., Baden-Baden 2017, § 34 Rn. 77d.

72 Neumann (Fn. 71), § 34 Rn. 74; krit. Weigend, Notstandsrecht (Fn. 62), 601.

73 Schuster, Notstandsalgorithmen beim autonomen Fahrzeug, RAW 2017, 13 (17); Weigend, Notstandsrecht (Fn. 62), 603 f.

abstrakten Merkmalen zu entscheiden, wer in einer Dilemma-Situation getötet werden soll; der Fahrer sehe hingegen die Personen vor sich und habe seine Entscheidung angesichts der konkreten Umstände zu treffen. Damit sei derjenige gerechtfertigt, der die höherrangige Pflicht wähle, d.h. quantitativ mehr Menschen rette als opfere. Dieser Differenzierung steht jedoch entgegen, dass die Fahrfunktion dann, wenn sie die Dilemma-Situation „erkennt“, vor derselben Entscheidung steht: Fortsetzung der Fahrt ohne Lenkimpuls – also untätig bleiben – oder Wahl des Ausweichweges – also aktiv werden.

Drittens wird in den Fällen des quantitativen Überwiegens ein übergesetzlicher Notstand erwogen. Die Rechtsprechung hat diese Rechtsfigur zwar noch nicht allgemein anerkannt,⁷⁴ mit der Überlegung, dass das geringere Übel gewählt werden darf,⁷⁵ wird aber im Schrifttum die Berufung auf einen übergesetzlichen Notstand teilweise zugelassen.⁷⁶ Hiergegen wird eingewandt, in einer Rechtsordnung, die den Einzelnen respektiere, sei niemand verpflichtet, aus Solidarität sein Leben zugunsten ihm unbekannter Menschen zu opfern. Zudem stoße ein Algorithmus, der von Nützlichkeitsabwägungen geprägt sei und „automatisch“ Menschenleben opfere, auf grundlegende Bedenken. Daher wird ein Ausweichen überwiegend als nicht zu entschuldigend und damit als strafbar bewertet.⁷⁷

Die eingesetzte Ethik-Kommission hatte demgegenüber angenommen, eine „allgemeine Programmierung auf eine Minderung der Zahl von Personenschäden“ könne „vertretbar“ sein.⁷⁸ Dies harmoniert damit, dass eine von MIT-Forschern bereits im Jahr 2015 durchgeführte weltweite Umfrage gezeigt hatte, dass für die Mehrheit der befragten Menschen die Anzahl der geretteten Leben im Vordergrund steht.⁷⁹ Diese utilitaristische Einstellung hat die von anderen Forschern entwickelte Online-Plattform „The Moral Machine“,⁸⁰ die Nutzer im Internet Dilemma-Entscheidungen treffen lässt, bestätigt. Für die Veröffentlichung der Ergebnisse im Jahr 2018 waren im-

74 Siehe BGH NJW 1953, 513; NJW 1989, 912 (913).

75 Neumann (Fn. 71), § 35 Rn. 54, 59 m.w.N.

76 Engländer, Das selbstfahrende Kraftfahrzeug (Fn. 56), 616; Hörnle/Wohlers, Trolley Problem (Fn. 63), 29.

77 Hilgendorf, Automatisiertes Fahren (Fn. 28), 805 m.w.N.

78 Bundesministerium für Verkehr und digitale Infrastruktur, Bericht (Fn. 42), S. 11, Regel Nr. 9.

79 J.-F. Bonnefon/A. Shariff/I. Rahwan, The Social Dilemma of Autonomous Vehicles, *Science* 2016, 1573.

80 <https://www.moralmachine.net/hl/de>.

merhin 40 Millionen Entscheidungen von Menschen aus 233 Ländern und Territorien ausgewertet worden.⁸¹

Stützen ließe sich eine entsprechende Programmierung mit weiteren Überlegungen. Konsequentialistisch betrachtet erscheint es unangemessen, denjenigen zu bestrafen, der eine möglichst große Zahl von Menschen retten will.⁸² Umgekehrt dürfte aus einer deontologischen Perspektive eine auf die Minimierung der Opferzahl ausgelegte Programmierung im Interesse eines jeden Einzelnen sein, da sie abstrakt im Vorfeld das Risiko jedes Einzelnen reduziert, der jemals konkret in eine solche Situation kommen sollte.⁸³ Auch kontraktualistisch lässt sich annehmen, dass eine (Zwangs-)Regel gerechtfertigt sein kann, sobald sie abstrakt im Interesse jedes Einzelnen ist.⁸⁴ Aus kantischer Perspektive wird hierbei niemand als „bloßes Mittel“ behandelt, vielmehr ist jeder ist auch „Zweck der Regelung“.⁸⁵ Im Übrigen könnte daraus, dass das System generell auf die „Schadensreduzierung“ ausgelegt sein muss (§ 1e Abs. 2 Nr. 2 lit. a StVG), geschlossen werden, dass eine Programmierung, die mehreren Menschenleben den Vorzug gibt, ohne hierbei qualitativ zu gewichten, nicht untersagt ist.

Daher erscheint die Annahme eines übergesetzlichen entschuldigenden Notstands in entsprechenden Fallkonstellationen vertretbar, sodass dem Täter kein strafrechtlicher Vorwurf gemacht werden sollte, wenn er eine Fahrfunktion einsetzt, die auf die Rettung möglichst vieler Menschen programmiert ist.

b) Fälle kumulativer Gefährdung

Denkbar sind weiter Situationen, in denen die Opfer kumulativ gefährdet sind. Im Schrifttum wurde diesbezüglich der Fall gebildet, dass in einer engen Straße drei Kinder auf die Straße springen und ohne Lenkimpuls alle drei sterben müssten.⁸⁶ Wird nach rechts gelenkt, werden zwei Kinder sterben, beim Lenken nach links findet nur ein Kind den Tod. Erachtet man ein Umsteuern für unzulässig, wären alle drei Leben verloren. Auch hier ist

81 *E. Awad et. al.*, The Moral Machine experiment, *Nature* 2018, 59.

82 *Hörnle/Wohlers*, Trolley Problem (Fn. 63), 30.

83 *Hevelke/Nida-Rümelin*, Selbstfahrende Autos (Fn. 55), 11 f.

84 *Hörnle/Wohlers*, Trolley Problem (Fn. 63), 31.

85 *Hevelke/Nida-Rümelin*, Selbstfahrende Autos (Fn. 55), 12; *Schuster*, Notstandsalgorithmen (Fn. 73), 18.

86 *Hilgendorf*, Automatisiertes Fahren (Fn. 28), 807.

eine Rechtfertigung ausgeschlossen, da jedes menschliche Leben denselben hohen Wert hat.⁸⁷ In derartigen Situationen dürften aber wohl die meisten Menschen versuchen, so viele Kinder wie möglich zu retten. Für die Auswahl des geringsten Übels sollte daher kein strafrechtlicher Vorwurf erhoben werden,⁸⁸ mithin eine entsprechende Programmierung nicht als unzulässig bewertet werden.

3. Tötung von Insassen

Die vorherigen Dilemma-Situationen sind dahingehend abwandelbar, dass das Leben der Insassen gefährdet ist. Denkbar ist, dass sich im Fahrweg eine Person befindet, der nur ausgewichen werden kann, indem das Kfz gegen eine Betonmauer gesteuert wird, wodurch der Insasse stirbt. Ebenso können sich sowohl im Fahrweg als auch im Kfz mehrere Personen befinden. Fraglich ist, ob eine Strafbarkeit bestehen kann, wenn ein autonomes Fahrzeug Insassen opfert, um eine Menschengruppe zu retten.

Verbreitet wird die Lösung erneut auf der Rechtfertigungsebene gesucht. Einerseits wird eine Parallele zum defensiven Notstand gezogen. Die Tötung eines Menschen, der eine Lebensgefahr für andere darstellt, sei gerechtfertigt, daher dürften auch Personen im Fahrweg überfahren und damit getötet werden.⁸⁹ Dies gelte umso mehr, wenn die Personen selbstverschuldet (z.B. aufgrund von Trunkenheit) in den Fahrweg geraten seien. Gegen eine Rechtfertigung spricht jedoch erneut, dass auch in den Fällen, in denen das Opfer eine Gefahrenquelle darstellt, Leben gegen Leben steht und der Umstand, dass ein Mensch eine Gefahr darstellt, noch kein Recht begründet, ihn zu töten. Daher hat die Rspr. z.B. im Haustyranen-Fall eine Rechtfertigung aus § 34 StGB versagt.⁹⁰ Im Hinblick auf § 216 StGB (Tötung auf Verlangen) muss eine Rechtfertigung sogar dann ausscheiden, wenn ein potentielles Opfer signalisieren würde, es solle überfahren werden, damit ein anderer Mensch gerettet wird (z.B. eine Mutter, die sich für ihr Kind aufopfern möchte).⁹¹ Schließlich wird sich ex ante kaum feststellen lassen, ob sich eine Person verschuldet oder unverschuldet im Fahrweg befunden hat.⁹²

87 A.A. Hörnle/Wohlers, Trolley Problem (Fn. 63), 18 f.

88 Hilgendorf, Dilemma-Probleme (Fn. 41), 696.

89 Weigend, Notstandsrecht (Fn. 62), 602 m.w.N.

90 BGHSt 48, 255.

91 A.A. Mitsch, Probleme (Fn. 61), 72 f.

92 Mitsch, Probleme (Fn. 61), 72; Weigend, Notstandsrecht (Fn. 62), 602.

Andererseits wird auch hier verfochten, dass zwei Unterlassungspflichten kollidieren und die Konstellation der rechtfertigenden Pflichtenkollision vorliege. Soweit nur ein Leben gegen ein Leben steht, soll die Lösung dem Zufallsgenerator überlassen werden, im Übrigen sei der Weg zu wählen, bei dem mehr Menschen gerettet werden.⁹³ Hiergegen ist wiederum einzuwenden, dass auch in diesen Fällen eine Handlungs- mit einer (gewichtigeren) Unterlassungspflicht kollidiert und eine Verrechnung von Menschenleben ausscheidet, weil jedes einzelne menschliche Leben höchsten Wert hat.

Schließlich wird die Lösung erneut auf der Schuldebene gesucht. Wenn ein menschlicher Fahrer in einer gegenwärtigen, nicht anders abwendbaren Gefahr für sein Leben einen Menschen tötet, um eine Gefahr von sich abzuwenden, handele er gemäß § 35 Abs. 1 StGB ohne Schuld.⁹⁴ Wenn der Fahrer eines Kfz eine entsprechend programmierte Fahrfunktion aktiviere und nutze, könne er sich ebenfalls auf diesen Entschuldigungsgrund berufen. Für den Halter eines autonomen Fahrzeugs, der nicht selbst in Gefahr schwebt, komme dagegen nur ein übergesetzlicher entschuldigender Notstand in Betracht, da die Fahrfunktion im Interesse der Passagiere aktiviert und genutzt werde. Diese Lösung harmoniert mit der Auffassung,⁹⁵ wonach die Fahrfunktion so zu programmieren ist, „dass die Fahrzeuginsassen unter allen Umständen geschützt werden“; es sei nicht geboten, das herannahende Fahrzeug zu vernichten (etwa durch Zusteuern auf einen Betonpfeiler oder durch einen Selbstzerstörungsmechanismus), selbst wenn die Zahl der Geretteten die Zahl der Geopferten übersteigt.

Gegen diese Lösung wird eingewandt, die Passagiere könnten nicht erwarten, dass ihnen das selbstfahrende Fahrzeug „neben anderen Bequemlichkeiten auch eine absolute Lebensversicherung für Kollisionsfälle bietet“; in Wahrheit stünden ökonomische Gründe im Vordergrund, da ein eingebauter Selbstzerstörungsmechanismus „kein besonders verkaufsförderndes Ausstattungsdetail“ sei.⁹⁶ Dieser Argumentation ist jedoch entgegenzuhalten, dass ein Passagier dann, wenn er selbst fahren würde, gem. § 35 Abs. 1 StGB ohne Schuld handeln würde. Der Umstand, dass eine Fahrfunktion in seinem Interesse genutzt wird, ändert an dieser grundsätzlichen Wertung nichts. Der Halter ist daher nicht verpflichtet, das Leben der Passagiere aufzuopfern. Im Übrigen stellt es ein dem Straßenverkehr immanentes allgemeines Risiko dar,

93 Weigend, Notstandsrecht (Fn. 62), 603.

94 Schuster, Notstandsalgorithmen (Fn. 73), 15.

95 Hilgendorf, Autonomes Fahren, in: Hilgendorf (Hrsg.), Autonome Systeme (Fn. 57), S. 143 (170).

96 Weigend, Notstandsrecht (Fn. 62), 604.

dass derjenige, der – aus welchen Gründen auch immer – sich auf der Fahrbahn befindet, Opfer eines Verkehrsunfalls werden kann. Daher darf auch der Hersteller nicht verpflichtet werden, dieses Risiko auszuschließen, indem er den Tod der Passagiere bzw. die Selbsterstörung des Kfz programmiert. Dies gilt selbst dann, wenn durch den Tod des einzigen Passagiers zahlreiche Menschen gerettet werden könnten.

4. Verletzungswahrscheinlichkeiten

In den bisher betrachteten Fällen betrug die Wahrscheinlichkeit der Tötung 100 %. Im Alltag wird aber z.B. ein junger Mensch einem Kfz eher ausweichen können als ein Hochbetagter. Zudem sind die Insassen durch Fahrgastzelle, Gurtstraffer und Airbags wesentlich besser geschützt als ein Fußgänger. In der Praxis differieren daher Art und Schwere der Verletzungen und damit die Überlebenschancen ganz erheblich.⁹⁷ Die heutigen Systeme können dies nicht berücksichtigen. Angesichts des technischen Fortschritts ist jedoch zu erwarten, dass eine autonome Fahrfunktion auch diese Faktoren einbezieht und risikominimierende Entscheidungen schneller und zuverlässiger als ein Mensch treffen wird. Die Berücksichtigung der unterschiedlichen Verletzungswahrscheinlichkeiten ist auch deshalb geboten, weil das Gesetz dem menschlichen Leben höchste Priorität einräumt. Soweit aber hieraus geschlossen wird, dass künftig besonders gefährdete Verkehrsteilnehmer (wie Fußgänger) stark geschützt und besonders sichere Fahrzeuge häufiger in Unfälle verwickelt werden,⁹⁸ geht dies zu weit. Wie dargelegt ist ein Hersteller nicht zu einer Programmierung verpflichtet, die das Leben der Passagiere opfert. Die höhere Überlebenschance ist nur dann von Bedeutung, soweit es um die Gefährdung von Menschen geht, die sich im Fahrweg und auf den möglichen Ausweichrouten befinden.

F. Fazit

Deutschland hat als erster Staat einen Rechtsrahmen sowohl für die technische Genehmigung als auch für den Betrieb autonomer Fahrzeuge im Regelbetrieb geschaffen. Die strafrechtliche Verantwortlichkeit von Fahrern,

97 Hevelke/Nida-Rümelin, Selbstfahrende Autos (Fn. 55), 13; Hilgendorf, Automatisiertes Fahren (Fn. 28), 807.

98 Hevelke/Nida-Rümelin, Selbstfahrende Autos (Fn. 55), 18.

Haltern, Passagieren, Herstellern und Kfz-Betrieben lässt sich mit der vorhandenen Dogmatik bestimmen. Bei Dilemma-Situationen hat der Schutz menschlichen Lebens höchste Priorität, wobei eine qualitative Gewichtung ausgeschlossen ist. Dagegen sollte demjenigen, der eine Fahrfunktion aktiviert und nutzt, die einer quantitativen Gewichtung folgt, kein strafrechtlicher Vorwurf gemacht werden. Befinden sich Menschen im Fahrweg, ist eine Programmierung zulässig, welche die Insassen absolut schützt. Befinden sich mehrere Menschen im Fahrweg und auf den möglichen Ausweichrouten, wird mit dem wachsendem technischen Fortschritt die Berücksichtigung von Art und Schwere möglicher Verletzungen und damit eine Auswahl nach der höheren Überlebenschance möglich werden.

Kurzbiografien der Autorinnen und Autoren

Nina Eckertz

ist Bachelor of Arts im Fach Medienwissenschaft (Medieninformatik) und Lehrassistentin am Institut für Digital Humanities der Universität zu Köln.

Lehrveranstaltungen und Vorträge im Bereich der Anwendung von künstlicher Intelligenz und Machine Learning im Bereich der Digital Humanities.

Schwerpunkt aktueller Projekte ist der Einsatz von Hyperspectral Imaging und die Datenvisualisierung für die Cultural Heritage Forschung.

Øyvind Eide

ist Professor für Digital Humanities – Historisch-Kulturwissenschaftliche Informationsverarbeitung und Vice Director, Center for Data and Simulation Science an der Universität zu Köln.

Seine Forschungsinteressen fokussieren sich auf transformative digitale Intermediale-Studien und theoretische Studien zur Modellierung in den Geisteswissenschaften und darüber hinaus.

Stephan Hobe

ist Professor für öffentliches Recht, Völker- und Europarecht, sowie Direktor des Instituts für Luftrecht, Weltraumrecht und Cyberrecht an der Universität zu Köln.

Lehrbücher im Völkerrecht, Europarecht und Weltraumrecht.

Schwerpunkt der Forschung sind Völkerrechtstheorie und neue Herausforderungen des Völkerrechts durch Cyberwaffen, Europäische Integration und die Fortentwicklung des Weltraumrechts.

Amina Hoppe

ist wissenschaftliche Mitarbeiterin am Institut für Kriminologie der Universität zu Köln.

Derzeitige Forschungsschwerpunkte im Strafrecht und der Kriminologie des digitalen Raums.

Christian Katzenmeier

ist Professor für Bürgerliches Recht und Zivilprozessrecht, Direktor des Instituts für Medizinrecht der Universität zu Köln und stellv. Direktor des

Cologne Center for ethics, rights, economics, and social sciences of health. Er ist u.a. Schriftleiter der Zeitschrift „Medizinrecht“ (MedR), Herausgeber der „Kölner Schriften zum Medizinrecht“, Mitherausgeber des „Heidelberger Kommentar Arztrecht Krankenhausrecht Medizinrecht“ (HK-AM) und des Laufs/Katzenmeier/Lipp, „Arztrecht“, Kommentator des Behandlungsvertrags (BeckOK) und des Rechts der unerlaubten Handlungen (NK-BGB und Handbuch der Beweislast). Seine Forschungsschwerpunkte sind das Medizinrecht, das Arztrecht, das Haftungs- und Versicherungsrecht sowie das Zivilprozessrecht.

Torsten Körber

ist Professor für Bürgerliches Recht, Kartell- und Regulierungsrecht, Recht der digitalen Wirtschaft und Direktor des Instituts für Energiewirtschaftsrecht (EWIR) der Universität zu Köln.

Er ist u.a. Mitherausgeber der Kommentare Immenga/Mestmäcker, Wettbewerbsrecht, und Säcker/Körber, TKG und TTDSG, sowie der Neuen Zeitschrift für Kartellrecht (NZKart).

Seine Forschungsschwerpunkte sind das Kartellrecht (in seiner ganzen Breite), zivil- und wettbewerbsrechtliche Aspekte des Energierechts und Telekommunikationsrechts sowie das Recht der digitalen Daten- und Plattformwirtschaft.

Axel Ockenfels

ist Wirtschaftswissenschaftler an der Universität zu Köln. Nach einer Gastprofessur an der Stanford University leitet er seit 2015 das Kölner „Exzellenzzentrum für Soziales und Ökonomisches Verhalten“. Er erforscht das menschliche Verhalten und entwickelt Designlösungen, wenn Märkte und Entscheidungsarchitekturen versagen oder Verhaltensänderungen erforderlich sind.

Er ist Mitglied im Wissenschaftlichen Beirat beim Bundeswirtschaftsministerium und in der Leopoldina, der Berlin-Brandenburgischen und weiteren Akademien der Wissenschaft tätig. In seiner Forschung kombiniert er Spieltheorie und Marktdesign mit der Verhaltensforschung und leistet sowohl Beiträge zu den Grundlagen der Verhaltenswissenschaft und Verhaltensmodellierung als auch zur Gestaltung realer Märkte und zu Anreizmechanismen.

Martin Schwamborn

ist Akademischer Rat a. Z. und Habilitand am Institut für Luftrecht, Welt-
raumrecht und Cyberrecht sowie am Lehrstuhl für Völker- und Europarecht,
Europäisches und Internationales Wirtschaftsrecht der Universität zu Köln.

Er wurde in Köln mit der Arbeit „Maßstäbe der europäischen Integration:
Möglichkeiten und Grenzen eines maßstabsorientierten Kooperationsver-
hältnisses zwischen BVerfG und EuGH“ promoviert. Die Arbeit wurde
mit dem Promotionspreis 2022 der Rechtswissenschaftlichen Fakultät ausge-
zeichnet.

Seine Forschungsschwerpunkte umfassen neben dem allgemeinen Völ-
ker- und Europarecht insbesondere das europäische und internationale
Wirtschaftsrecht sowie aktuelle Herausforderungen des Luft-, Weltraum-
und Cyberrechts.

Martin Paul Waßmer

ist Professor für Strafrecht und Strafprozessrecht, geschäftsführender Direk-
tor des Instituts für Strafrecht und Strafprozessrecht an der Universität zu
Köln sowie Studiendekan der Rechtswissenschaftlichen Fakultät.

Er ist Autor zahlreicher Kommentierungen zum deutschen, europäischen
und internationalen (Wirtschafts-)Strafrecht, Herausgeber mehrerer Schrif-
tenreihen und Sammelbände, Verfasser einer strafrechtlichen Klausurenlehre
und eines Lehrbuchs zum Medizinstrafrecht.

Schwerpunkte seiner Forschung bilden das Wirtschafts- und Steuer-
strafrecht, Medizin-, Ausländer-, Computer- und Internetstrafrecht sowie
das Ordnungswidrigkeitenrecht.