Künstliche Intelligenz und neue Verantwortungsarchitektur*

Timo Rademacher

Die Frage nach der *Architektur* von etwas lädt zu Überblicken ein, fordert auf zur Analyse nicht nur einzelner Bestandteile des betrachteten Etwas, sondern verlangt auch nach Betrachtung und Bewertung des Zusammenwirkens dieser einzelnen Bestandteile.[1] Sich aus dieser Perspektive *Künstlicher Intelligenz und Verantwortung* zu nähern, ist nicht ungefährlich. Angesichts der Vielgestaltigkeit, Komplexität und Kompliziertheit des Themas „KI" läuft ein solches Unternehmen Gefahr, zur kleinteiligen Beschreibung eben doch von Einzelbestandteilen zu werden, die den erhobenen Vollständigkeitsanspruch am Ende gleichwohl enttäuschen müsste. Der vorliegende Beitrag muss und darf sich daher von vornherein beschränken und bescheidenere Ziele setzen: Unter I. soll zunächst versucht werden, so gedrängt wie möglich darzustellen, was unter der Überschrift *Künstliche Intelligenz und Verantwortung* vielleicht schon als konzediert gelten darf. Danach – unter II. – werden Schlaglichter geworfen auf drei ausgewählte Aspekte, wo Künstliche Intelligenz (KI) uns dazu auffordert oder zumindest dazu einlädt, unsere Begriffe und Vorstellungen von Verantwortung neu, vielleicht sogar disruptiv zu denken. Der Beitrag schließt unter III. mit einem als Appell formulierten Fazit.

I. Konzediertes

1. Neue Herausforderungen?

Zunächst also zum Konzedierten, wobei sich hier sogleich die Frage aufdrängt: Ist in der hier zu reflektierenden Diskussion denn überhaupt

* Der Verfasser dankt *Jens-Peter Schneider* sowie *Nikolaus Marsch* für zahlreiche Diskussionen und Anregungen zum Thema.

1 Siehe nur *Hoffmann-Riem*, AöR 142 (2017), 1 (8ff.), zu der mit dem Architekturbegriff ermöglichten „Ausweitung des Blickes" auf die „Einbettung [algorithmischer Systeme] in komplexe Infrastrukturen und damit mittelbar [auf] die Bedeutung weiterer Funktionsvoraussetzungen und Verwendungsmöglichkeiten der Informations- und Kommunikationstechnologie".

(schon) etwas konzediert? Eine gewisse Einigkeit scheint immerhin bezüglich der Herausforderungen zu bestehen, die „KI“[2] bereithält:[3] Wir haben es zu tun mit der „Ausbreitung“ von Akteuren, oder, in der besser vom Menschen abgrenzenden Terminologie von *Bruno Latour*: mit der Ausbreitung von *Aktanten*,[4] die unsere virtuelle oder physische Realität oder die heranwachsende Mischform aus beidem (Stichwort „on*life*“[5]) verändern können, und dabei folgende vier Eigenschaften aufweisen:

- Digitale Agenten sind *erstens* zunehmend allgegenwärtig[6] und grenzüberschreitend[7] aktiv und aktiviert;
- *zweitens* ist ihnen durch die technische Komplexität und Kompliziertheit ihrer Entscheidungsfindungsprozesse eine bestimmte Form von Unberechenbarkeit und situativer Unerklärbarkeit ihrer Handlungen zu eigen; die Stichworte hierzu lauten *black-box*-Phänomen[8] und „Autonomierisiko“;[9]
- ferner sind die digitalen Agenten potentiell stark vernetzt, d. h. treten nicht nur als isolierte Systeme auf, sondern können untereinander oder auch „hybrid“ in Form von Verbindungen mit Menschen verknüpft sein, sodass sie über Hintergrund-„Wissen“ verfügen und aufgrund von „Wissen“ nach außen handeln, das das Alter Ego im Moment der Interaktion nicht hat und auch *ex post* nur schwer zurückverfolgen kann.[10] In der Literatur etablieren sich hier der Begriff „Vernetzungsrisiko“

2 Zur Definition von KI, die im Rahmen dieses Beitrags nicht vertieft werden kann, siehe v. a. *Hochrangige Expertengruppe für Künstliche Intelligenz*, Eine Definition der KI: Wichtigste Fähigkeiten und Wissenschaftsgebiete, abrufbar unter https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60664 <27.3.2020>.
3 Vgl. *Teubner*, AcP 218 (2018), 155; *Fateh-Moghadam*, ZSTW 2019, 863 (875f.).
4 *Latour*, Das Parlament der Dinge: Für eine politisch Ökologie, 2001, S. 93ff.
5 *Hildebrandt*, The Modern Law Review 79 (2016), 1 (4f.).
6 *Rademacher*, JZ 74 (2019), 702 (706).
7 *Hoffmann-Riem*, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 1 Rn. 46ff.; zu den Möglichkeiten einer Re-Territorialisierung insbes. des Internets siehe v. a. *Svantesson*, Solving the Internet jurisdiction puzzle, 2017.
8 Maßstäbe setzend *Wischmeyer*, AöR 143 (2018), 1, bes. S. 42ff.; nun auch monographisch ausgearbeitet und aktualisiert von *Martini*, Blackbox Algorithmus, 2019, passim, m. w. N.
9 *Teubner* (Fn. 3), S. 163ff.; *Cornelius*, ZIS 2020, 51 (53f.).
10 Siehe *Spiecker gen. Döhmann*, in: Fehling/Schliesky (Hrsg.), Neue Macht- und Verantwortungsstrukturen in der digitalen Welt, 2016, S. 53 (60f.).

bzw. – für die hybriden Mensch-Maschine-Netzwerke – der Begriff „Verbundrisiko";[11]

- und schließlich, *viertens*, weisen digitale Agenten aufgrund ihrer in der Regel vergangenheitsbasierten Lernmethoden[12] starke Pfadabhängigkeiten auf und projizieren diese auf ihre Nutzerinnen und Nutzer potentiell mit Wirkungen für die Gegenwart und Zukunft zurück.

Dass um diese vier Herausforderungen herum eine Verantwortungsarchitektur gebaut werden soll, auch hinsichtlich dieses Ziels besteht Einigkeit,[13] jedenfalls in den westlichen Diskursräumen.

2. Neue Architektur – aus bewährten Bausteinen

Nun sind die genannten Herausforderungen abstrakt und *je für sich* betrachtet keineswegs unbekannt. Der Umgang mit *black boxes* ist – als Aufgabe – vertraut, sei es mit Blick auf Unternehmen oder Tiere, sei es mit Blick auf Menschen;[14] der Umgang mit Risikotechnologien ist – als Aufgabe – ebenso vertraut, das Umwelt- und das Chemikalienrecht liefern Vorbilder, die in der KI-Diskussion auch schon breit herangezogen werden;[15] auch der Umgang mit unklaren Schadensverursachungen und Beweisproblemen und – besonders aus dem europäischen Verwaltungsrecht – mit analogen oder elektronischen Vernetzungen ist – als Aufgabe – vertraut.[16]

---

11 *Teubner* (Fn. 3), S. 201ff. bzw. 196ff.; aufgreifend *Zech,* Risiken digitaler Systeme, Weizenbau Series #2, 2020, S. 48ff. bzw. 26; *Cornelius* (Fn. 9), S. 54f.; *Fateh-Moghadam* (Fn. 3), S. 865f., 875; mit leicht abweichender Terminologie auch *Spiecker gen. Döhmann* (Fn. 10), S. 65ff.

12 *Rademacher*, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 225 Rn. 35 m. w. N.

13 Siehe nur die KI-Empfehlungen des Rates der OECD on Artificial Intelligence v. 22.5.2019, wo „responsible stewardship of trustworthy AI" als zentrales Ziel genannt ist; abrufbar unter https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 <27.3.2020>.

14 *Wischmeyer* (Fn. 8), S. 54.

15 *Martini* (Fn. 8), S. 113ff.

16 Siehe *Schneider*, in: Hill/Schliesky (Hrsg.), Herausforderung e-Government. E-Volution des Rechts- und Verwaltungssystems, 2009, S. 89, mit dem Hinweis, dass die dogmatischen Erkenntnisse aus analogen Vernetzungsstrukturen durchaus auf elektronische Vernetzungen übertragen werden können.

> „Dabei stellt die genaue Verteilung der Verantwortungslasten eine Herausforderung dar. Ein Bedarf für die Einführung gänzlich neuer Verantwortungsstrukturen lässt sich jedoch nicht erkennen".[17]

Das „Neue" liegt also vielfach nur in der (neuen) Rekombination von Bekanntem. Nun soll hier wie eingangs angekündigt gerade nicht der weitgehend aussichtslose Versuch unternommen werden, eine Gesamtarchitektur entwerfen zu wollen, welche die herausforderungsvolle „Verteilung der Verantwortung" im Einzelnen (siehe soeben) und die diffizilen, vornehmlich verwaltungs- und sozialwissenschaftlich zu erforschenden Wechselwirkungen der einzelnen Bausteine[18] zeichnet. Aber immerhin soll auf hoher Abstraktionsebene eine Taxonomie der bekannten und neu zusammenzusetzenden Verantwortungs-Bausteine benannt werden:

a. Bauverbote

Ein grundlegendes Verantwortungselement im Zusammenhang mit KI, das selten ausdrücklich genannt wird (wahrscheinlich, weil oft reflexhaft als fortschrittsfeindlich wahrgenommen), ist das „Bauverbot". Gemeint ist damit der Verzicht auf eine bestimmte Technologie oder bestimmte technologische Methode (selten),[19] das Verbot bestimmter Anwendungen (häufig, etwa aus der Genetik bekannt[20]), oder auch Moratorien, wie

17 *Wischmeyer* (Fn. 8), S. 36.

18 *Hoffmann-Riem*, Innovation und Recht – Recht und Innovation, 2016, § 6; *ders.* (Fn. 1), S. 8ff.

19 *Teubner* (Fn. 3), S. 175.

20 Vgl. etwa das zwar nicht KI-, aber immerhin Big-Data-spezifische Verwendungsverbot für gentechnische Untersuchungen in § 18 Gendiagnostikgesetz. Viele der in der Lit. mit KI-Systemen verbundenen Gefahrenanalysen basieren weniger auf technischen Defiziten als vielmehr ganz schlicht auf Datennutzungen, die als illegitim und übergriffig wahrgenommen werden, beispielhaft die Analyse von *Danaher*, in: Yeung/Lodge (Hrsg.), Algorithmic Regulation, 2019, S. 98 (106-109). Dass sich die regulative Kraft der Einwilligung in Datenverarbeitungen als weitgehend dysfunktional erwiesen hat, um Verarbeitungsszenarien so zu steuern, dass sie als hinreichend legitim empfunden werden, wurde bereits mehrfach nachgewiesen; vgl. dazu umfassend *Hermstrüwer*, Informationelle Selbstgefährdung, 2016.

jüngst von Seiten der EU-Kommission kurz angedacht für den Einsatz von Gesichtserkennungssoftware.[21]

b. Materiell-rechtliches Programm

Selbstverständlich geht es dann nicht ohne materiell-rechtliches Programm, sei es gesetzlich, sei es in Form vertraglicher oder deliktischer Verkehrs- und Sorgfaltspflichten beim Training und Einsatz von KI. Gesetzgeber,[22] Verwaltung und Gerichte,[23] aber auch die Industriestandards setzenden privaten Institutionen[24] sind hier noch in der Bringschuld. Stattdessen allein oder auch nur maßgeblich auf Selbstregulierung der privaten Akteure oder gar auf Marktmechanismen zu setzen, scheidet – wie *Wolfgang Hoffmann-Riem* stets betont hat – „angesichts der Machtasymmetrien" im IT-Bereich aus.[25] In der Sache klärungsbedürftig wird etwa sein, ob und wo algorithmische Entscheidungen akzeptiert werden können, die im Ergebnis und/oder im Allgemeinen gut funktionieren; es aber im Einzelfall wegen des erwähnten *black-box*-Phänomens nicht mehr menschlich nachvollziehbar ist, *warum* eine bestimmte Aktion so und nicht anders ausgeführt wurde.[26]

---

21 Vgl. aus der Berichterstattung *Fanta*, EU erwägt Verbot von Gesichtserkennung, Netzplotik.org v. 17.1.2020, abrufbar unter https://netzpolitik.org/2020/eu-erwaegt-verbot-von-gesichtserkennung/ <27.3.2020>.

22 Zum aktuellen Stand der Gesetzgebung auf EU-Ebene siehe das White Paper der Kommission *On Artificial Intelligence – A European approach to excellence and trust*, COM(2020) 65 final v. 19.2.2020, sowie die am selben Tag veröffentlichte *European Strategy for data*, COM(2020) 66 final.

23 Für einen von der Rspr. zu entwickelnden beweglichen Maßstab, der mit der zunehmenden Leistungsfähigkeit digitaler Agenten strenger wird, *Teubner* (Fn. 3), S. 194.

24 Siehe die weitgehend im Status *under development* befindlichen ISO/OEC-Standards zu KI, dazu mit Übersichten www.iso.org/committee/6794475.html <27.3.2020>.

25 Schon *Hoffmann-Riem* (Fn. 1), S. 39.

26 Man könnte diesen Punkt auch verfahrensrechtlich verstehen (was ich hier nicht tue, da „Begründung" z. B. im Sinn von § 39 VwVfG sich in der Forderung nach dem handlungs*bestimmenden* Grund erschöpft – und der könnte eben durchaus „because computer said so and got it right in 95% of test cases" lauten); ggf. zählt der Punkt dann zum nächsten Baustein *Institutionen*, *Verfahren*, *Verfahrensarrangements*. Zur Frage der Begründbarkeit bei/trotz/durch KI in diesem Band → *Wischmeyer*, bes. bei S. 78ff., 81ff. Jüngst ausf. hierzu *Watson/Floridi*, https://ssrn.com/abstract=3509737 <17.3.2020>.

c. Institutionen, Verfahren, Verfahrensarrangements

Ein zentrales Feld der bisherigen öffentlich-rechtlichen Diskussion bildet dann die Frage nach Institutionen, Verfahren und Verfahrensarrangements, die den Einsatz von KI vorbereiten, begleiten und *ex post* evaluieren und ggf. korrigieren können.[27] Zu diesem sehr weiten Feld seien an dieser Stelle nur einige Bemerkungen gesetzt: Zwar scheint der immer wieder geforderte zentrale Algorithmen-TÜV[28] oder die zentrale KI-Agentur[29] derzeit noch nicht im Aufbau begriffen. Doch zeigen sektorale Lösungen wie die geplanten umfassenden Auskunftsrechte der Landesmedienanstalten, die der neue Medienstaatsvertrag gegenüber Medienplattformen und Medienintermediären einführt,[30] dass die Gesetzgeber die Forderungen nach solchen tiefgehenden Kontrollstrukturen durchaus aufgreifen. Richtigerweise gehört hierher aber nicht nur die Frage nach punktuellen Auskunftspflichten. Genauso wichtig ist die Schaffung einer Sub-Architektur zur breiten Wissensgenerierung *vor* dem akuten Aufsichts- oder Kontrollfall.[31] Ziel muss es sein, das derzeit oft nur bei privaten Akteuren vorhandene

27 Siehe bes. *Tutt*, Administrative Law Review 68 (2017), 83ff.; in diesem Band → *Wischmeyer*, bei S. 82f., Fn. 33ff. m. w. N. Diskussionsstränge zusammenführend *Wischmeyer*, in: ders./Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 75 Rn. 36ff., 44ff., m. w. N.

28 Bes. wirkmächtig der frühe Appell von *Tutt* (Fn. 27), bes. S. 105ff.

29 Zur Erforderlichkeit einer Digitalagentur www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/bericht-der-kommission-wettbewerbsrecht-4-0.html <27.3.2020>; nicht Aufsicht und Kontrolle, sondern der Wirtschafts- und Forschungsförderung dient die KI-Agentur Bayerns, vgl. den Nachweis unter https://zentrum-digitalisierung.bayern/aufbau-der-bayerischen-ki-agentur/ <27.3.2020>.

30 Siehe § 86 (für Medienplattformen und Benutzeroberflächen), § 95 (für Medienintermediäre), jeweils i. V. m. § 56 (Auskunftsrechte und Ermittlungsbefugnisse) des Entwurfes eines Staatsvertrags zur Modernisierung der Medienordnung in Deutschland v. 5.12.2019.

31 Diese könnten z. B. vergaberechtlicher Natur sein (vgl. hierzu schon *Wischmeyer*, in: Goldhammer/Kulick (Hrsg.), Der Terrorist als Feind? Personalisierung im Polizei- und Völkerrecht, 2020, S. 193 (211f., 213)) oder auf „hybrides" Personal bei den privaten IT-Dienstleistern setzen (dazu *Reiling*, Der Hybride, 2015, bes. S. 171ff.). Natürlich käme auch in Betracht, vermehrt auf Eigenentwicklungen staatlicher (Aufsichts-)Behörden zu setzen, was aber zumindest praktisch nur in begrenztem Umfang realistisch erscheint. *Trennscharfe* rechtliche Maßgaben, wann der Staat auf interne Wissensreservoirs zurückgreifen muss, und wann er externe Ressourcen nutzen darf, gibt es freilich nicht, vgl. schon *Groß*, in: Röhl (Hrsg.), Die Verwaltung/Beiheft 9: Wissen – Zur kognitiven Dimension des Rechts, 2010, S. 135 (152).

Wissen frühzeitig auf die kontrollierenden Institutionen zu übertragen.[32] Nur dann können diese ihre Steuerungs-, Kontroll- und Gewährleistungsverantwortung auch praktisch wahrnehmen. Daneben scheint es mir auch eine im weiteren Sinn verfahrensrechtliche Frage zu sein, ob der Einsatz von autonomen Softwareagenten gegenüber denjenigen, die mit ihnen interagieren, stets offengelegt werden muss. Eine solche Kennzeichnungspflicht wird vielfach gefordert,[33] existiert – in vielleicht unterschätzter Breite – auch bereits im geltenden Datenschutzrecht,[34] und macht es wie kaum ein anderer Verantwortungsbaustein erforderlich, bereits im Alltag, also jenseits pathologischer Situationen, eine Abgrenzbarkeit von und in Mensch-Maschine-Systemen zu erhalten.[35]

d. Haftung

Ein weiterer, jüngst viel diskutierter Baustein ist das Haftungsrecht.[36] Auch hier werden bekannte Instrumente zum Umgang mit Nicht-Wissen und dem Handeln unter Arbeitsteilung wie Beweislastumkehr, Kausalitäts-

---

32 Hierzu jüngst BVerfG, NVwZ 2019, 52 Rn. 24: „Der Gesetzgeber mag […] kurzfristig darauf vertrauen können, dass sich fachliche Wissenslücken durch Erkenntnisfortschritte in Fachkreisen und Wissenschaft schließen. Längerfristig dürfte der Gesetzgeber dem jedoch nicht tatenlos zusehen, weil er sich so seiner inhaltlichen Entscheidungsverantwortung entzieht, privatem Fachwissen ungesteuert weitreichenden Einfluss auf staatliche Entscheidungen eröffnet und eine einheitliche Rechtsanwendung nicht gewährleistet ist. Der Gesetzgeber muss dann, sofern die fachlichen Zusammenhänge weiter ungeklärt sind, für eine zumindest untergesetzliche Maßstabsbildung beispielsweise durch Einsetzung fachkundiger Gremien zur Festlegung einheitlicher Maßstäbe und Methoden sorgen oder wenigstens genauere Regeln für die behördliche Entscheidung zwischen mehreren vertretbaren Auffassungen vorgeben." Der Fall betraf immissionsschutzrechtlich relevante Wissensdefizite.

33 *Wischmeyer* (Fn. 8), S. 20.

34 Vgl. Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g, Art. 15 Abs. 1 lit. h, Art. 22 Abs. 3 i. V. m. ErwGr 71 Abs. 4 DSGVO; insofern zum aktuellen Stand der Diskussion *Kumkar/Roth-Isigkeit*, JZ 75 (2020), 277, bes. S. 283ff. zu den offenen Fragen.

35 Zu den in der Lit. verbreiteten Zweifeln daran, dass eine solche Abgrenzung (stets) gelingen kann bzw. gelingen muss, siehe oben, bei Fn. 10f., und unten, bei Fn. 89, bzw. unten, Fn. 85.

36 Siehe zur europ. Übersicht den Bericht der *Expert Group on Liability and New Technologies*, Liability for Artificial Intelligence, 2019, sowie aus der Lit. bes. *Teubner* (Fn. 3), S. 155ff.; ferner *Foerster*, ZfPW 2019, 418 (430ff.); *Borges*, NJW 2018, 977 (980ff.); *Denga*, CR 2018, 69; *Jakl*, MMR 2019, 711 (713ff.); *Linardatos*, ZIP 2019, 504ff.; jeweils m. w. N. Speziell zur Haftung für APR-Verletzungen

vermutung, Gehilfenhaftung, Gefährdungshaftung, Gesamtschuld mit Innenregress und haftungsakzessorische Pflicht-Versicherungsmodelle aufgerufen. Der kürzlich von *Herbert Zech* unterbreitete Vorschlag einer KI-Versicherung nach dem Vorbild der gesetzlichen Unfallversicherungen[37] erscheint dabei besonders charmant. Denn neben der Haftung als solcher könnten sich gerade auch die besonderen, auf Wissensgenerierung gerichteten Strukturen der Unfallversicherungen als vorbildhaft erweisen.[38]

e. Sanktionen

Der letzte große Baustein betrifft das Sanktionenrecht. Besonders passend erscheint hier der englische Begriff der *Blamability*:[39] Es geht darum, für bestimmte als besonders gravierend wahrgenommene Fehlleistungen – ganz salopp formuliert – noch jemanden „anschreien" zu können. Vornehmer und rechtstechnischer ausgedrückt geht es vor allem um Fahrlässigkeitsdelikte, auf die sich der Fokus der strafrechtlichen KI-Diskussion richtet.[40] Die Problematik, dass individuelle Verantwortungsbeiträge angesichts der Vernetzung von KI-Systemen bzw. ihrer Verbindung zu Mensch-Maschine-Hybriden immer diffuser werden, ist auch hier keine unbekannte, sondern lässt sich – nach Einschätzungen in der strafrechtlichen Literatur – unter Rückgriff auf die Grundsätze der Verantwortungsabgrenzung im Rahmen arbeitsteiliger Prozesse bewältigen.[41]

---

durch KI-Agenten *Oster*, UFITA 82 (2018), 14ff. Spezifisch *staats*haftungsrechtlich vertiefende Diskussionsbeiträge fehlen – soweit ersichtlich – bislang.

37 *Zech*, in: Deutscher Juristentag (Hrsg.), Verhandlungen des 73. Deutschen Juristentages Hamburg 2020 / Bonn 2022, Bd. I: Gutachten, 2020, A 105ff.

38 Siehe grundlegend zu dieser Aufgabe § 14 Abs. 1 SGB VII.

39 Zum – allerdings vergleichsweise selten gebrauchten – Begriff etwa *Sankowski*, Journal of Aesthetic Education 22 (1988), 49ff. Zur Passung überkommener philosophischer *Blamability*-Konzepte auf KI auch noch unten, um/bei Fn. 79.

40 Ausf. *Fateh-Moghadam* (Fn. 3), S. 875ff., bes. 883ff.; *Cornelius* (Fn. 9), S. 59f.; *Yuan*, RW 2018, 477 (493f.); *Gaede*, Künstliche Intelligenz – Rechte und Strafen für Roboter? 2019, S. 81ff.; Überlegungen zu Veränderungen der strafrechtlichen Verantwortlichkeit auch über die Fahrlässigkeitsdelikte hinaus bei *Beck*, ZIS 2020, 41, bes. S. 43ff. Ein ausf. Kommentierung von § 42 BDSG findet sich – zur Vorgängernorm im BDSG a. F. – bei *Golla,* Die Straf- und Bußgeldtatbestände der Datenschutzgesetze als Teil des Schutzes des informationellen Selbstbestimmungsrechts, 2015.

41 Siehe *Beck* (Fn. 40), S. 43; *Yuan* (Fn. 40), S. 493f., m. w. N.

Es lässt sich allerdings auch KI-spezifisches *Vorsatz*-Strafrecht finden: § 42 BDSG[42] in Verbindung mit Art. 22 DSGVO.[43] Ohne zu tief eintauchen zu können, sei dieser Normkomplex hier eigens erwähnt, weil die Normen (auch) von der strafrechtlichen Diskussion – soweit ersichtlich – bisher weitgehend ignoriert werden;[44] sie teilen damit das Schicksal zahlreicher Regelungen des Nebenstrafrechts. Dabei ist § 42 BDSG für die *KI-und-Verantwortung*-Diskussion potentiell hoch relevant. Denn in Kombination mit Art. 22 DSGVO bewehrt § 42 Abs. 2 BDSG viele der soeben genannten KI-Verantwortungsbausteine bereits jetzt strafrechtlich, rundet also sozusagen die DSGVO sanktionsrechtlich ab, soweit sie jetzt schon KI-bezogene Regelungen enthält.[45] Das Problem dabei ist, dass die Norm gerade in Verbindung mit dem seinerseits extrem unbestimmten Art. 22 DSGVO[46] nicht bestimmt genug sein dürfte, um einer verfassungsgerichtlichen Nachprüfung standhalten zu können.

---

42 § 42 Abs. 2 BDSG lautet, soweit hier von Interesse: „Mit Freiheitsstrafe bis zu zwei Jahren oder mit Geldstrafe wird bestraft, wer personenbezogene Daten, die nicht allgemein zugänglich sind, 1. ohne hierzu berechtigt zu sein, verarbeitet […] und hierbei gegen Entgelt oder in der Absicht handelt, sich oder einen anderen zu bereichern oder einen anderen zu schädigen."

43 Abs. 1 dieser Norm lautet: „Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt."

44 Im KI-spezifischen strafrechtlichen Schrifttum findet § 42 BDSG bislang, soweit ersichtlich, keine Beachtung, vgl. z. B. die ansonsten sehr ausführlichen Überlegungen *de lege lata* und *ferenda* bei *Fateh-Moghadam* (Fn. 3), S. 876ff. Von einem „theoretisch vermutlich relativ großen Anwendungsfeld" geht hingegen aus BDSG-HK-*Heghmanns,* 2020, § 42 Rn. 36, das allerdings v. a. durch das Strafantragserfordernis des Abs. 3 S. 1 praktisch doch relativ klein ausfallen könnte.

45 Das in Art. 22 DSGVO enthaltene grds. Verbot vollständig automatisierter Entscheidungen im Einzelfall, einschließlich Profiling, kann als zentrale KI-spezifische Norm der DSGVO gelten. Das Verbot, geregelt in Abs. 1, gilt aber nicht absolut, sondern weicht unter den – sehr weiten – Voraussetzungen des Abs. 2 einer Erlaubnis, die ihrerseits aber wiederum unter dem Vorbehalt „angemessener Maßnahmen zur Wahrung der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person" steht (so Abs. 2 lit. b, ähnlich Abs. 3). Bei der Liste dieser „angemessenen Maßnahmen", zu denen nach Art. 22 Abs. 4 DSGVO schon *de lege lata* auch ausdrückliche Diskriminierungsverbote zählen, finden sich zahlreiche der soeben genannten Verantwortungsbausteine bereits jetzt wieder, vgl. statt vieler *Scholz*, in: Simitis/Hornung/Spiecker, Datenschutzrecht, 2019, Art. 22 DSGVO Rn. 56-64 m. w. N.

46 Zu den vielen Unzulänglichkeiten von Art. 22 DSGVO, aber auch zum Potential der Regelung, statt vieler *Bygrave*, in: Yeung/Lodge (Hrsg.), Algorithmic Regulation, 2019, S. 248ff.

3. … aber mit neuen Leistungsmaßstäben!

Schließlich scheint mir mittlerweile ein weiterer Punkt konzediert: Es spricht viel dafür, an KI im Rahmen der soeben genannten Bausteine zumindest partiell *höhere* Anforderungen zu stellen als an Menschen.[47] Das gilt schon allein deshalb, weil von den Entscheidungen „eines" zentral eingesetzten digitalen Agenten im Zweifel mehr Personen betroffen sein werden als vom Handeln eines menschlichen Entscheiders.[48] Selbst wenn also beispielsweise die automatisierte Gesichtserkennung fast gleich gut funktionieren sollte wie die menschliche,[49] so folgt daraus noch nicht, dass sie nun nach gleichen Maßstäben wie die Streifenpolizistin eingesetzt werden darf.

Aber auch jenseits der damit angesprochenen Skaleneffekte eines Computerprogramms und der damit verbundenen Gefahren[50] dürfte *Armin Nassehi* mit seiner Beobachtung recht haben: In Zukunft „dürfte das ausgezeichnete Privileg des Menschen darin bestehen, nicht nur Fehler zu machen, sondern Fehler machen zu dürfen".[51]

II. Disruptives: Wessen Verantwortung, und wofür?

Der zweite Teil des Beitrags gilt nun drei ausgewählten Aspekten im Zusammenhang von *KI und Verantwortung*, die sicherlich nicht mehr als konzediert gelten können, die aber – so zumindest meine Annahme – besonders relevant und dringend sind.

1. Maschinenverantwortung?

Die Bausteine, die im ersten Teil dieses Beitrags aufgezählt sind, sind – quasi phänomenologisch – das Ergebnis von Recherchen, was in der Literatur unter den Suchworten „Verantwortung" und „KI" an rechtlichen,

---

47 Vgl. etwa *Teubner* (Fn. 3), S. 194 m. w. N.
48 *Danaher* (Fn. 20), S. 110-112.
49 *Simonite*, MIT Technology Review v. 17.3.2014, abrufbar unter www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/ <27.3.2020>.
50 Siehe für den bes. sensiblen polizeilichen Kontext *Rademacher*, AöR 142 (2017), 366 (374, 376).
51 *Nassehi,* Muster: Theorie der digitalen Gesellschaft, 2. Aufl. 2019, S. 226.

rechtsdogmatischen und rechtspolitischen Vorschlägen zu finden ist. Die Diskussionen oder Streitstände betreffen hier meist die Details der Grenzen des Auslegbaren, grundrechtlich Zumutbaren oder Praktikablen. Anders ist es, wenn man gezielt fragt, *wer* eigentlich Verantwortung für die Entscheidungen von autonomen Agenten übernehmen soll.

Klassisch und auf hoher Abstraktionsebene lässt sich Verantwortung, frei nach *Jan Henrik Klement*, als *Antworten* verstehen, oder, etwas genauer: als Antworten-*Können* und Danach-*Handeln*-Können.[52] Und als *Klement* sein Buch „Verantwortung" schrieb, war auch noch ganz klar, dass der so Adressierte, also das Verantwortungssubjekt, ein Mensch ist.[53]

Das bleibt heute nicht mehr unwidersprochen: Einen vielbeachteten Vorschlag für eine neue Adressierung von Verantwortung machte 2018 *Gunter Teubner* im *AcP* unter dem Titel *Digitale Rechtssubjekte*? *Teubner* arbeitet mit dem folgenden Verantwortungs-Begriff:

> „Verantwortung im strengen Sinn ist das Einstehenmüssen für Entscheidung unter Ungewissheit, deren Resultat nicht prognostizierbar ist.[54] […]"

Auf aktuelle und vor allem auf sich abzeichnende Formen von KI übertragen folgt für *Teubner* daraus dann das entscheidend Neue:

> „Wenn Computer Entscheidungen unter Ungewissheit treffen, dann ermöglicht dies, etwas gänzlich Neues zu entdecken. […] Ja, Verantwortung gewinnt ihre eigentliche Bedeutung erst hier: Als Einstehenmüssen für den ‚Sprung ins Dunkle'.[55] […] Diesen Sprung ins Dunkle […] nicht nur Menschen anzuvertrauen, sondern Algorithmen zu überlassen, darin besteht das fundamental Neue".[56]

Denn andernfalls würden

> „Softwareagenten [unausweichlich] Verantwortungslücken [verursachen], da ihr autonomes Handeln einen massiven Kontrollverlust menschlicher Akteure mit sich bringt. […] Die Dynamik der Digitali-

52 *Klement*, Verantwortung, 2006, S. 576. Siehe für die Adaption der Definition speziell für autonome Systeme *Schulz*, Verantwortlichkeit bei autonom agierenden Systemen, 2015, S. 39ff.

53 Bes. deutlich *Spiecker* (Fn. 10), S. 60: „Spricht man über Verantwortung mit rechtswissenschaftlichem Bezug, dann ist in einer freiheitlichen Gesellschaft immer der Einzelne, das Individuum, adressiert."

54 *Teubner* (Fn. 3), S. 175.

55 Ebd., S. 176.

56 Ebd.

> sierung erzeugt unaufhörlich verantwortungsfreie Räume, die sich in Zukunft ausweiten werden"

– es sei denn, wir stufen Softwareagenten „selbst als verantwortliche Akteure" ein, so *Teubners* zentrale These.[57] Die wiederum findet eine Stütze in einem systemtheoretisch informierten Analogieschluss zum Unternehmen, zur Organisation, der KI-Systeme in vielen Aspekten gleichen würden.[58]

Diese Bereitschaft zum Wagnis hat etwas verlockend zukunftsgewandtes. Die Gleichsetzung von digitalen Agenten mit Unternehmen oder mit Menschen ist deshalb so charmant, weil dann, wenn diese Gleichsetzung funktionieren würde, eine Blaupause auch für eine Verantwortungsarchitektur in einem gesamtheitlichen Sinne existieren würde. Beim Weiterlesen des zitierten Beitrags wird aber rasch genug klar, dass sich sein Autor zu öffentlich-rechtlichen, strafrechtlichen oder auch nur zivilrechtlichen Verantwortungsbausteinen *insgesamt* gar nicht verhalten, sein Vorbild „Unternehmen" gar nicht *insgesamt* heranziehen kann oder auch nur will. Mit der Unternehmensanalogie ist also keine Architektur im anspruchsvollen Sinn beschrieben, es wird keine *Verantwortungs-Gesamtrechnung* aufgemacht. Vielmehr bezieht *Teubner* seinen Verantwortungsbegriff – zwar nicht in der Definition, wohl aber in den anschließenden Detail-Ausführungen – ganz auf vertragsrechtliche Zurechnung und vertragliche bzw. deliktische Haftung für das Agieren digitaler Agenten. Es geht dann in der „Pointe"[59] ganz klassisch-zivilrechtlich um die Anwendung von Stellvertretungsregeln oder um die Gehilfenhaftung nach § 278 BGB oder § 831 BGB, natürlich jeweils analog.

Die Analogien lassen sich weitertreiben:[60] Besonders interessant erscheint die Orientierung von KI-Agenten am Vorbild von Tieren,[61] was

---

57 Ebd., S. 157.

58 Zu der damit z. B. von *Teubner* (Fn. 3), S. 155 begründeten Notwendigkeit der Anerkennung einer Teilrechtsfähigkeit digitaler Agenten aber treffend *Marietta Auer*, Verfassungsblog v. 30.9.2019, aufgerufen unter https://verfassungsblog.de/rechtsfaehige-softwareagenten-ein-erfrischender-anachronismus <13.2.2020>: „Denn aus der bloßen Eigenschaft, eine kommunizierende Kollektividentität zu sein, folgt noch keine Notwendigkeit, als juristische Person oder sonstige rechtsfähige Entität anerkannt zu werden."

59 *Fateh-Moghadam* (Fn. 3), S. 877.

60 Ausf. Überlegungen aus anglo-amerikanischer Perspektive bei *van den Hoven van Genderen*, in: Barfield/Pagallo (Hrsg.), Research handbook on the law of artificial intelligence, 2018, S. 213 (216ff.), m. w. N.

61 So z. B. *Vöneky*, OdW 2020, 9 (12) m. w. N.

dann zur analogen Anwendung der Tierhalterhaftung nach § 833 BGB oder in den USA zur Anwendung der Regeln für Drogenspürhunde auf *predictive-policing*-Systeme führen könnte.[62] Populär ist auch die Frage, ob KI nach dem Vorbild von Risikotechnologien, wie der Atomkraft oder der Genetik, reguliert werden kann und soll.[63] Vielleicht noch gängiger ist es nur, den Menschen gleich selbst als KI-Vorbild und KI-Maßstab heranzuziehen.[64] Zumindest semantisch weisen in diese Richtung die vielen Stellungnamen und Arbeitsgruppen, die von „vertrauenswürdiger KI" statt vom vertrauenswürdigen Umgang mit KI[65] oder ohne menschliche Mittelung[66] von „verantwortlicher KI" sprechen.[67] Ganz ähnlich verhält es sich mit dem 2019 veröffentlichten Aufruf von 23 Computerwissenschaftlern, die neue akademische Disziplin der *Machine Bevaviour Studies* zu gründen,[68] was auch universitär die „Verflüssigung der Grenzziehung von Person, Tier und Maschine"[69] unterstreichen würde.

Hier zeigt sich aber die Gefahr der Argumentation mit analogen Vorbildern: Die Orientierung an ihnen ist sinnvoll, solange sie einen „Steinbruch" bilden, aus dem sich die Liste der oben genannten Bausteine für eine KI-Verantwortungsarchitektur ergeben soll. Als „Blaupausen"[70] für ganze Architekturen sind die analogen Vorbilder aber gefährlich: Denn wir knüpfen an die Begriffe, die sich in der Diskussion bekanntlich leicht

---

62 Mit diesen Überlegungen schon *Rich*, University of Pennsylvania Law Review 164 (2016), 871 (913ff.), und auch *Ferguson*, William & Mary Law Review 55 (2014), 1283 (1358ff.).

63 Bes. gründlich *Martini* (Fn. 8), S. 113-155.

64 So etwa *Linardatos* (Fn. 36), S. 507; *Matthias*, zitiert nach *Ziemann*, in: Hilgendorf/Günther (Hrsg.), Robotik und Gesetzgebung, 2013, S. 183 (183f.).

65 So die Ethik-Leitlinien für eine vertrauenswürdige KI der *Hochrangigen Expertengruppe für Künstliche Intelligenz*, mit letztem Stand veröffentlicht am 8.4.2019, abrufbar unter https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top <27.3.2020>; zur Kritik an der Wortwahl *Markschies*, Warum sollte man einem Computer vertrauen? F.A.Z. v. 23.2.2019, S. 11.

66 Siehe in diesem Sinne die KI-Empfehlungen des Rates der OECD on Artificial Intelligence v. 22.5.2019, wo von „responsible *stewardship* of trustworthy (sic!) AI" die Rede ist, Hervorhebung hier. Das Dokument ist abrufbar unter https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 <27.3.2020>.

67 So etwa das Forschungsprojekt *Responsible Artificial Intelligence* an der Universität Freiburg i. Br., siehe www.frias.uni-freiburg.de/de/foerderprogramme/schwerpunkte/responsible-ai <27.3.2020>.

68 *Rahwan et al.*, Nature 568 (2019), 477.

69 *Fateh-Moghadam* (Fn. 3), S. 869.

70 So *Martini* (Fn. 8), S. 113ff., der freilich – richtiger Weise – stets gesondert und kritisch die „Analogiefähigkeit" der von ihm angedachten Blaupausen untersucht.

verselbstständigen und von den differenzierenden Ausführungen im – meist – zweiten Teil rechtswissenschaftlicher Aufsätze ablösen, holistische Verantwortungs- oder eben auch *Nicht*-Verantwortungszuschreibungen, die sich im Laufe der Rechtsentwicklung etabliert haben – wie zum Beispiel beim Tier, jedenfalls nach Überwindung des mittelalterlichen Tierstrafrechts.[71] Dann aber droht zumindest eine Verwirrung der Diskussion.[72] Zu sagen „Ein digitaler Agent müsse wie ein Unternehmen Verantwortung für seine Entscheidungen übernehmen", ist misslich, wenn eigentlich nur die Übernahme von Verantwortung spezifisch *in der Form von monetärer Haftung gemeint ist*.

Dabei ist zu beachten, dass sich viele der vorhin genannten Verantwortungsbausteine nach der ihnen eigenen Funktionslogik *durchaus* durch digitale Agenten selbst ausfüllen lassen, jedenfalls weitgehend:

- Im Bereich der monetären Haftung dürfte es nur ein praktisches Problem sein, z. B. Sondervermögen als Haftungsmassen oder eine KI-Unfallversicherung zu schaffen, wie *Zech* es mit guten Argumenten vorschlägt.[73]
- Auch die Grundfunktion von Verantwortung, nämlich das Antworten und nachvollziehbar Rechenschaft über das Warum einer Entscheidung ablegen können, ist ein Verantwortungsbaustein, der unter den Schlagworten *explainable artificial intelligence* (xAI) oder *interpretable machine learning* (iML)[74] bereits automatisiert wird bzw. werden soll.[75]
- Das institutionalisierte Testen und Evaluieren von KI-Agenten ist eine Verantwortungs-Funktion, die sich in Zukunft *ohne* die Hilfe eben von Test- und Evaluierungs-Algorithmen angesichts der wachsenden Kom-

71 *Ziemann* (Fn. 64), S. 185ff.

72 Vgl., noch deutlicher zu den überschießenden Wirkungen von Analogien, *Schirmer*, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 123 Rn. 24: „[T]he trap snaps shut."

73 Siehe oben, Fn. 37; dort auch zahlreiche weitere Nachw. zu alternativen Haftungsmodellen.

74 Zu den Begriffen *Watson/Floridi,* https://ssrn.com/abstract=3509737 <17.3.2020>.

75 Siehe → *Wischmeyer*, in diesem Band, S. 82, Fn. 32. Zum aktuellen Forschungsstand *Garnelo/Shanahan*, Current Opinion in Behavioral Sciences 29 (2019), 17ff.; *Samek/Montavon/Vedaldi/Hansen/Müller*, Explainable AI, 2020. Dazu, dass auch xAI-Anwendungen den sog. *automation bias* (= ungerechtfertigtes Vertrauen in die Richtigkeit computergestützter Entscheidungen seitens der NutzerInnen) auslösen können, *Heaven*, MIT Technology Review v. 29.1.2020, abrufbar unter https://www.technologyreview.com/2020/01/29/304857/why-asking-an-ai-to-explain-itself-can-make-things-worse/ <20.4.2020>, m. w. N.

plexität der Systeme und der Arbeit im Umfeld von Geschäftsgeheimnissen kaum mehr denken lässt[76] – Stichwort *compliance by design*.[77]

- Schließlich lässt sich selbst die Erstellung des materiell-rechtlichen Programms, dem Software-Agenten unterliegen, *jenseits* von Wertentscheidungen, also wo Recht eine reine Optimierungsaufgabe hat, irgendwann eventuell auf KI übertragen.[78]

Es gibt aber mindestens einen Baustein, der sich nicht auf KI rejustieren lässt: Der moralische Vorwurf, die *Blamability*,[79] oder, rechtlich(er) ausgedrückt, das Bedürfnis nach *strafrechtlicher* – oder, was ich hier aber nicht vertiefen will: demokratischer[80] – Verantwortungszuschreibung. Im englischen Diskurs wird dieser Aspekt sprachlich deutlicher: Software-Agenten können, in der Terminologie von *Luciano Floridi*, *accountable* für ihre Handlungen sein; sie können aber nicht in einem vollständigen Sinne *responsible* sein, schlicht weil sie konzeptionell nicht sinnvoll bestrafbar sind.[81] Das idealerweise normstabilisierende Unwerturteil, das in der Kri-

---

76 Statt vieler *Kroll et al.*, University of Pennsylvania Law Review 165 (2017), 633 (660ff.).

77 Speziell zur Kontrolle von KI durch KI-gestützte *impossibility structures* auch *Cornelius* (Fn. 9), S. 57; zum Begriff der *impossibility structures Rademacher* (Fn. 6), S. 703f, sowie unten, bei Fn. 123.

78 Zu der Gefahr, dass auch an sich (noch) nicht für eine automatisierte Gesetzesanwendung geeignete Materien für computergestützte Verarbeitung – und d. h. wahrscheinlich: bewusst vereinfachend – geregelt werden könnten, *Unger*, in: ders./Ungern-Sternberg (Hrsg.), Demokratie und künstliche Intelligenz, 2019, S. 113 (121), und dort bes. die Nachw. in Fn. 58.

79 Zum Begriff oben, bei Fn. 39.

80 Die Möglichkeit der Abwahl in demokratischen Herrschaftsstrukturen, die ebenfalls eine Vorwurfskomponente transportiert bzw. transportieren soll, lasse ich im Folgenden außer Betracht; zu Formen „digitaler Herrschaft als demokratischem Zurechnungsproblem" mit zahlreichen weiteren Nachw. *Unger* (Fn. 78), bes. S. 118ff.

81 Siehe *Floridi*, in: Anderson/Leigh Anderson (Hrsg.), Machine Ethics, 2011, S. 184 (201-203, 205): „Recall that moral accountability is a necessary but insufficient condition for moral responsibility." Aus dem strafrechtlichen Schrifttum ebenso z. B. *Gless/Silverman/Weigend*, New Criminal Law Review 19 (2016), 412 (415ff.). Ob sich hieran etwas ändert, falls KI-Agenten einmal Gefühle haben oder simulieren können sollten (vgl. *van den Hoven van Genderen* (Fn. 60), S. 238; *Kuehn/Haddadin*, IEEE Robotics and Automation Letters (RA-L), 2 (2016), 72ff.) und wir dann beginnen, ihnen Bewusstsein und eventuell sogar Rechte zuzuschreiben, soll hier offen bleiben. Mit diesen Zukunftsüberlegungen ausf. *Gaede* (Fn. 40), S. 57ff., 66ff., der allerdings selbst für diesen Fall daran zweifelt, dass es ein geeignetes, gegenüber Robotern wirksames Strafübel im überkommenen Sinn gibt (S. 66f.).

minalstrafe enthalten ist,[82] ist ihnen gleich und ist damit wirkungslos.[83] Natürlich kann man den sanktionsrechtlichen Verantwortungs-Baustein für überflüssig halten,[84] wie es namentlich *Floridi* vorschwebt:

> „We can stop the regress of looking for the responsible individual when something evil happens, because we are now ready to acknowledge that sometimes the moral source of evil or good can be different from an individual or group of humans".[85]

Ich halte *Blamability* als Teil einer Verantwortungs-Gesamtrechnung für unverzichtbar.[86] Falls das stimmt, dann erfordert das – entgegen der be-

---

82 *Fateh-Moghadam* (Fn. 3), S. 877ff., mit weiteren ausf. Überlegungen zur Dysfunktionalität eines anthropomorphisierenden „Roboterstrafrechts".

83 Ähnlich *Fateh-Moghadam* (Fn. 3), S. 876ff., dort, Fn. 76, auch zur teilweise vertretenen Gegenauffassung. Wenn demgegenüber in der philosophischen Diskussion argumentiert wird, dass *blamability* sich vom *punishment* dahingehend unterscheide, dass ein Adressat von *blame* dadurch zwar häufig einen Nachteil (*punishment*) empfinde, das aber nicht zu den Essentialia von *blame*/*blamability* gehöre, wird dem hier – jedenfalls in Übertragung auf KI – nicht gefolgt: Ein Vorwurf (*blame*), der strukturell, d. h. strukturbedingt stets und notwendig, auf emotionale Gleichgültigkeit beim Adressaten stößt, verfehlt die damit vom und beim Sender intendierte Befriedigungswirkung zwangsläufig, sodass jedenfalls nicht von *sinnvoller blamability* gesprochen werden kann. Anders, freilich aus Prä-KI-Tagen stammend, z. B. *Smith*, Proceedings of the Aristotelian Society, New Series 109 (2009), 31 (55); ähnlich wohl auch *Sankowski* (Fn. 39), S. 51, 59 (mit der Frage, ob fiktionale literarische Charaktere sinnvoll *blamable* sind; *Sankowski* beantwortet die Frage positiv).

84 Vgl. hierzu auch *Burchard*, Normative Orders Working Paper 2/2019 – Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft, 2019, S. 24ff., der darauf hinweist, dass unter einer wohlfahrtsstaatlich-sicherheitsrechtlichen Strafrechtstheorie (i. S. v.: „Verunmöglichung oder zumindest die substantielle Minimierung von Rechtsgutsverletzung") die „Verheißungen von KI" im Falle ihrer Realisierung den Zweck des Strafrechts erfüllen und es damit gleichsam überflüssig machen könnten.

85 *Floridi* (Fn. 81), S. 210.

86 In diesem Sinne auch *Beck* (Fn. 40), S. 48: „Die dargestellten Veränderungen [= Diffusion von Verantwortung durch Digitalisierung] könnten zur Folge haben, dass in der digitalen Welt […] kaum noch Strafen verhängt werden. Das kann jedoch zu Problemen führen, wenn die Gesellschaft durch die fehlende Übernahme persönlicher Verantwortung beunruhigt und die Normgeltung bezweifelt wird. […] Die Zuschreibung individueller Verantwortung des Staates zum Bürger bzw. der Bürger untereinander ist grundlegend für unsere gegenseitige Wahrnehmung sowie unsere Selbstwahrnehmung." Ferner ebendort, S. 49: „Insofern [wäre] der umfassende Verzicht auf das Strafrecht nicht unproblematisch." Vgl. auch, aus dem zivilrechtlichen Schrifttum, *Schirmer* (Fn. 72), Rn. 11.

kannten systemtheoretisch fundierten Annahme, das „soziale Substrat [einer Organisation] sei [mehr als] eine Vielheit konkreter Menschen“[87] – doch wieder eine Zergliederung der system-theoretisch zusammengefügten kommunikativen Entscheidungsketten in einzelne Entscheidungsträger. Und wenn man *dann* nur auf digitale Agenten als „eigenverantwortliche“ Entscheidungsträger stößt,[88] dann steht die Verantwortungslücke, die durch die Analogie zum Unternehmen geschlossen schien, doch wieder offen. Das gesellschaftliche Akzeptanzreservoir, das auch und gerade im Bedürfnis bzw. in der Fähigkeit zum nachträglichen Schuldvorwurf liegt, kann eben nicht gehoben werden gegenüber einer „anonymen Matrix“[89] aus miteinander verwobenen digitalen Aktanten und menschlichen Akteuren mit angehängter Haftungs- oder Versicherungsmasse, sondern nur gegenüber individualisierten Menschen.[90] Die strafrechtswissenschaftliche Diskussion ist deshalb darum bemüht, vor allem im Rahmen der Fahrlässigkeitsdelikte *menschliche* Verantwortungsbeiträge beim Inverkehrbringen und Nutzen von KI-Agenten identifizierbar zu halten.[91]

Aufbauend auf der soeben referierten Einschätzung sei hier das Plädoyer erlaubt, dass wir als Rechtswissenschaft insgesamt vorsichtig sein sollten, analoge Vorbilder als Blaupausen für Verantwortungsarchitekturen – also als Vorbilder mit einem scheinbaren gesamtheitlichen Anspruch – auszuweisen, ohne dabei zugleich auch eine *Verantwortungs-Gesamtrechnung* aufzumachen, die dann alle drei großen Rechtsgebiete mindestens mit bedenkt.

2. Bürgerverantwortung?

Das nächste Schlaglicht – Bürgerverantwortung? – soll zunächst mit einem ausländischen Beispiel eingeleitet werden: Finnland zwingt seine Bürger bzw. genauer gesagt: deren Mobilitätsdienstleister seit 2018, unter ande-

87 *Teubner* (Fn. 3), S. 164.

88 Das wäre m. E. auch der Unterschied zum Unternehmensstrafrecht, wie es in anderen Staaten existiert; vgl. mit diesem Einwand gegen die hier vertretene Auffassung schon *Simmler/Markwalder*, ZStW 129 (2017), 20 (44); wie hier *Wohlers*, BJM 2016, 113 (123f.).

89 Ebd., S. 202.

90 Jedenfalls nach der noch g. h. M. in der strafrechtlichen Lit., Nachw. bei *Fateh-Moghadam* (Fn. 3), S. 876ff.; u. bestimmten U. aufgeschlossener aber z. B. *Cornelius* (Fn. 9), S. 60ff.

91 Dazu schon die Nachw. oben, in Fn. 40.

rem Bewegungs- und Zahlungsdaten mit anderen Mobilitätsdienstleistern zu teilen – im Interesse eines reibungsloseren, plattform- und KI-gestützten intermodalen Verkehrs der Zukunft.[92] Das Stichwort hierfür lautet: *Mobility as a Service*. Das finnische Gesetz ist derzeit freilich so konzipiert, dass der Austausch aggregierter Daten ausreicht, jedenfalls soweit es um die besonders sensiblen Standortdaten geht. Der europäische Rechtsrahmen, der für diese Verkehrs-Anwendungen gilt, lässt aber bereits jetzt zu, dass auch nationale Vorschriften zum zwingenden Austausch personenbezogener Daten gemacht werden.[93] *Data sharing* wird quasi zur Bürgerpflicht, zu einer neuen Verantwortung des einzelnen Datensubjekts, durch Nicht-Mehr-Zurückhaltung „seiner" Daten zum IT- und KI-Fortschritt beizutragen.

In diese Richtung weisen auch die neue Datenstrategie der EU[94] und das *White Paper* zur geplanten KI-Regulierung, die beide am 19. Februar 2020 vorgestellt wurden. So heißt es im *White Paper* zu den *Gefahren* künstlicher Intelligenz:

> „For example, as a result of a flaw in the object recognition technology, an autonomous car can wrongly identify an object on the road and cause an *accident* […]. [T]hese risks can […] *be related to problems with the availability and quality of data* or to other problems stemming from machine learning".[95]

Die noch unausgesprochene Konsequenz, die im hervorgehobenen Teil des Zitats verborgen ist, scheint eindeutig: Wer „seine" Daten nicht für die neuen EU-weit geplanten *data pools* zur Verfügung stellt,[96] der gefährdet

92 Vgl. dazu Part III, Chap. 1, Sect. 1, und Chap. 2 des finnischen *Act on Transportation Services*, in Kraft seit 1.7.2018.

93 Vgl. Art. 10 RL 2010/40/EU, Erwägungsgrund 5 der VO 217/1926.

94 COM(2020) 66 final, z. B. S. 5f.: "The infrastructures should support the creation of data pools enabling Big Data analytics and machine leaning, in a manner compliant with data protection legislation and competition law, allowing the emergence of data-driven ecosystems. […] Currently there is not enough data available for innovative re-use, including for the development of artificial intelligence."

95 COM(2020) 65 final, S. 12. Hervorhebung hier.

96 Vgl. EU-Kommission, A European strategy for data, COM(2020) 66 final, S. 14, 16ff., 21, 22f. Zur Diskussion aus kartellrechtlicher Perspektive *Schweitzer*, GRUR 2019, 569.

nicht nur den Fortschritt oder den reibungslosen Verkehr, sondern vielleicht, unter Umständen, je nach Sachbereich sogar Menschenleben.[97]

Fortgeschritten ist die Diskussion um individuelle Pflichten zum *data sharing* schon im Antidiskriminierungsrecht: Um zu verhindern, dass KI-Systeme diskriminierende Wirkungen haben,[98] bedarf es grundsätzlich Datensätze, die die diskriminierenden Eigenschaften enthalten, und zwar in möglichst repräsentativer Form.[99] Diese Erkenntnis läutet im Antidiskriminierungsrecht schon einen Paradigmenwechsel ein, weg vom Grundsatz des „Es geht niemanden etwas an, welche Hautfarbe ich habe" hin zu einem Grundsatz der Wissensgenerierung als Gemeinwohlaufgabe.[100] Nur noch kurz erwähnt sei hier schließlich das neue Digitale-Versorgung-Gesetz, welches das Bundesversicherungsamt im Interesse der Forschung zur zentralisierten Sammlung – Stichwort *data pools* – bestimmter Patientendaten berechtigt.[101] Damit sind wir schon sehr nah an den Schutz von Menschenleben durch *data sharing* herangerückt (zur aktuellen Corona-Krise siehe sogleich, bei Fn. 106).[102]

Meine These zu dieser neuen Bürgerverantwortung: *Data sharing* darf keine verhaltenssteuernden Konsequenzen haben, und darf deshalb auch grundsätzlich nicht mit verhaltenssteuernden Anreizen verknüpft werden. Dabei habe ich weniger Sorge vor einem potentiell gläsernen Bürger – ich denke nicht, dass der noch zu verhindern ist – oder Sorge um die Datensicherheit – wobei dies sicherlich das vordringlichste *technische* Forschungsfeld der nächsten Jahre sein dürfte. Die Dystopie, die mir realistischer er-

97 Die EU-Kommission (Fn. 96), S. 13 sieht „compulsory data access rights" als sektorspezifisch zu bestimmende Ausnahme an, vorranging bzw. allgemein soll mit Anreizen für einen Datenaustausch B2B gearbeitet werden.

98 Hierzu statt vieler *Hacker*, CMLR 55 (2018), 1143 (1146ff.).

99 *Kroll et al.* (Fn. 76), S. 685ff., dort auch, S. 688f., zu alternativen, datenschutzfreundlicheren Methoden wie „synthetischen" Trainingsdaten.

100 *Tischbirek*, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 103 (115ff.).

101 Einschlägig sind die novellierten §§ 303a ff. SGB V; dazu, auch zum Verhältnis zur DSGVO, *Kühling/Schildbach*, NZS 2020, 41. Zur Notwendigkeit/Sinnhaftigkeit des data sharing speziell im medizinischen Bereich siehe auch die Plädoyers von *Johnson*, BBC News v. 7.2.2014, abrufbar unter www.bbc.com/news/health-25988534 <27.3.2020> und – bes. nachdrücklich – *Moore-Colyer*, Silicon.co.uk v. 5.7.2017, abrufbar unter www.silicon.co.uk/e-regulation/data-sharing-privacy-216521 <27.3.2020>.

102 Bes. deutlich *Moore-Colyer* (Fn. 101).

scheint, ist das neue Ideal eines Gemeinwohl-optimierten Bürgers.[103] Die Vereinnahmung des Bürgers als Wissensgenerator oder, noch deutlicher: als Trainingsdatenlieferant für KI-Systeme liegt dabei sicherlich im Gemeinwohlinteresse und die EU-Kommission hat ohne Zweifel recht, dass es *ohne* eine solche Vereinnahmung keine gute KI geben wird. Es geht nicht ohne Daten, und auch wenn die Kommission vor allem auf Industriedaten abstellt: Es geht eben auch nicht ohne personenbezogene Daten.[104] Die Gefahr, an die Datenlieferung im nächsten Schritt quasi beiläufig gemeinwohlorientierte Verhaltensanreize zu setzen, folgt dem aber auf dem Fuß. Hier bedarf es m. E. früher und klarer Grenzen, die zu ziehen zum Beispiel im Bereich des Umweltschutzes angesichts der Herkulesaufgabe „Klimaerwärmung" extrem schwerfallen wird.

Die vorgenannten Überlegungen, das sei ergänzt, waren Teil des Manuskripts dieses Beitrags, bevor in der 12. Kalenderwoche 2020 wegen der rasanten Ausbreitung des SARS-CoV-2-Viruses („Corona") das öffentliche Leben in Europa zum Stillstand gebracht wurde. Damit stand die Frage – um im Bild zu bleiben – im gleißend hellen Schlaglicht, ob nicht ein massenhaftes, auf die frühzeitige Erkennung und Isolierung von individuellen Infizierten gerichtetes *data sharing* zwischen Telekom-Unternehmen, Kreditkartendienstleistern, Plattform- und Social-Media-Betreibern usw. hin zu den staatlichen Gesundheitsbehörden – zum Zwecke des sog. *contact trackings* bzw. *tracings* mit anschließenden, individuellen Folgemaßnahmen – ein alternatives Mittel zur Eindämmung der Pandemie hätte sein können.[105] Asiatische Staaten – darunter durchaus auch demokratisch-rechtsstaatlich organisierte wie Südkorea – hatten mit entsprechenden

---

103 Die KI-Diskussion kann hier an die bereits konsolidiertere *Nudging*-Debatte (dazu jüngst der Überblick von *Hufen*, JuS 2020, 193ff.) anschließen, z. B. unter dem Stichwort *hypernudging Yeung*, in: Yeung/Lodge (Hrsg.), Algorithmic Regulation, 2019, S. 21 (34f.), sowie *Danaher* (Fn. 20), S. 107f., der in der Kombination von verhaltenspsychologisch informierten *nudges* mit algorithmischer Personalisierung die größte Gefahr für menschliche Unabhängigkeit und damit Autonomie ausmacht. Literarisch ferner *Zeh*, Corpus Delicti, 2010.

104 Vgl. hierfür, wenn auch nicht in dieser Deutlichkeit, COM(2020) 65 final, S. 18f.

105 Siehe den – zurückgezogenen – Vorschlag für einen § 5 Abs. 12 IfSG v. 20.3.2020, Bearbeitungsstand 23:23 Uhr: „(12) [1]Für den Fall einer epidemischen Lage von nationaler Tragweite kann die zuständige Behörde zum Zwecke der Nachverfolgung von Kontaktpersonen technische Mittel einsetzen, um Kontaktpersonen von erkrankten Personen zu ermitteln, sofern aufgrund epidemiologischer Erkenntnisse gesichert ist, dass dies zum Schutz der Bevölkerung vor einer Gefährdung durch schwerwiegende übertragbare Krankheiten erforderlich ist. [2]Unter den Voraussetzungen nach Satz 1 kann die zuständige Behörde von jedem, der geschäftsmäßig Telekommunikationsdienste erbringt oder daran mit-

tracking- und *tracing*-Methoden zumindest anfangs gute Erfolge erzielt.[106] Das Thema ist im Schnittfeld von Forschungs-, Steuerungs- und Überwachungszwecken angesiedelt[107] und ihm sollte aufgrund der notstandsähnlichen Lage keine maßstabsbildende Funktion zugebilligt werden.[108] Auch geht es dem Grundrechte beschränkenden Staat hier nicht um die oben als Dystopie bezeichnete Optimierung des Einzelnen im Gemeinwohlinteres-

---

wirkt (Diensteanbieter) die Herausgabe der vorhandenen Telekommunikationsverkehrsdaten, der für die Ermittlung des Standortes eines Mobilfunkgerätes erforderlichen spezifischen Kennungen und die zur Durchführung von Maßnahmen nach Satz 4 erforderlichen Daten der möglichen Kontaktpersonen von erkrankten Personen verlangen. [3]Erforderlichkeit und Zweck der Maßnahme sind durch die zuständige Behörde zu dokumentieren. [4]Die zuständige Behörde kann die nach Satz 1 und 2 ermittelten Kontaktpersonen von dem Verdacht einer Erkrankung informieren. [5]Die zuständige Behörde darf zu diesem Zweck personenbezogene Daten verarbeiten. [6]Nach Beendigung der Maßnahmen ist die Löschung der Daten zu dokumentieren.", abrufbar unter https://fragdenstaat.de/dokumente/4075-anderung-des-infektionsschutzgesetzes-und-weiterer-gesetze/ <1.4.2020>. Verwiesen wird a.a.O., S. 21f., zur Rechtfertigung der angedachten Ermächtigung auf die positiven Erfahrungen des *trackings* in Südkorea. Richtig ist der Einwand des BfDI, dass ein solches Instrument angesichts der Eingriffstiefe einige Gewähr für seine Geeignetheit zur Zweckerreichung bieten muss, um eingesetzt werden zu dürfen, siehe Stellungnahme des BfDI v. 23.3.2020 zum Entwurf eines Gesetzes zum Schutz der Bevölkerung bei einer epidemischen Lage von nationaler Tragweite, S. 4. Für plausible Bedenken an der Geeignetheit der zunächst angedachten Regelung *Abeler/Bäcker/Buermeyer*, Netzpolitik.org v. 29.3.2020, abrufbar unter https://netzpolitik.org/2020/corona-tracking-datenschutz-kein-notwendiger-widerspruch/ <1.4.2020>. Dort auch der Vorschlag für ein freiwilliges und dank des Einsatzes der Bluetooth-Technologie datensparsames Alternativmodell zu den asiatischen Vorbildern.

106 Siehe mit knappen Berichten zu Singapur, Südkorea und Taiwan *M. Müller et al*., Wie Behörden ihre Bürger in der Corona-Krise mittels Smartphone-Daten überwachen, NZZ v. 14.3.2020, abrufbar unter www.nzz.ch/digital/coronavirus-wie-behoerden-die-buerger-in-der-krise-ueberwachen-ld.1546033?utm_source=pocket-newtab <27.3.2020>. Speziell Singapur scheint allerdings nicht nur auf die diskussionsbeherrschenden *tracking*-Apps gesetzt zu haben, sondern auf klassische Methoden der Durchsetzung von Hausarresten bzw. Quarantäne-Maßnahmen, siehe den Selbstbericht von *Hein*, F.A.Z. v. 26.3.2020, abrufbar unter https://www.faz.net/aktuell/wirtschaft/digitec/high-tech-gegen-coronavirus-kontrolle-in-singapur-16697321.html.

107 Zu letzterem auch noch unten, bei Fn. 123.

108 Vgl. zu den Gefahren der Argumentation in und mit Notstandslagen *Kingreen*, Whatever it Takes? Der demokratische Rechtsstaat in Zeiten von Corona, VerfBlog v. 20.3.2020, abrufbar unter https://verfassungsblog.de/whatever-it-takes/ <27.3.2020>: „Die krisentypische Einigkeits- und Entschlossenheitsrhetorik ist diskursfeindlich".

se, sondern um elementaren Lebens- und Systemschutz. Allerdings ist der Fall auch vorliegend von Interesse, illustriert er doch in höchstmöglicher Klarheit, wie wenig das seiner Natur nach abstrakte Datenschutzrecht[109] in konkreten Abwägungslagen mit akut bedrohten Rechtsgütern *jenseits prozeduraler Sicherungen* befriedigende *materielle* Leit- und damit auch Grenzlinien für den Einsatz von KI bzw. von neuen, datenintensiven Technologien generell vorgeben kann.[110] Es bleibt nur, diese materiellen Grenzen (wie in den vergangenen Jahren vielfach eingeübt) aus den abstrakten Missbrauchsgefahren von *data sharing* bzw. *tracking/tracing* zu gewinnen, also den Vergleich mit China zu bemühen, was angesichts der aktuellen Verfasstheit des deutschen demokratischen Rechtsstaats – vorsichtig ausgedrückt – hochspekulativ erscheint. Nur das grundrechtliche Wägen unter Heranziehung solcher Spekulationen kann dann zu der Einschätzung führen, dass generelle Ausgangs- und/oder Kontaktsperren, wie sie im März 2020 Realität wurden, gegenüber einem anlassbezogenen[111] *data sharing* bzw. *tracking* mit individuell zugeschnittenen Folgemaßnahmen einen *weniger* „tiefgreifenden Einschnitt" in Freiheits- und Bürgerrechte darstellen,[112] so dass das staatlich verordnete *tracking* zu unterbleiben habe.[113] Nach Überwindung der krisenhaften Lage muss nochmals mit Nachdruck gefragt werden, ob dem Schutz vor *hypothetisch-abstrakten* Gefahren staatli-

---

109 Im Fall des *trackings* von Corona-Verdachtsfällen einschlägig sind, vorbehaltlich der Verwendung ausschließlich anonymisierter/aggregierter Daten, Art. 6 Abs. 1 lit. c, d und/oder e, Art. 9 Abs. 2 lit. c, g und/oder h DSGVO, Art. 15 *ePrivacy*-Richtlinie 2002/58/EG.

110 Vgl. hierzu das klare *Statement on the processing of personal data in the context of the COVID-19 outbreak* v. 19.3.2020 des European Data Protection Board, abrufbar unter https://edpb.europa.eu/sites/edpb/files/files/news/edpb_statement_2020_processingpersonaldataandcovid-19_en.pdf <27.3.2020>, wo es auf S. 3 heißt: "Invasive measures, such as the 'tracking' of individuals (i.e. processing of historical non-anonymised location data) could be considered proportional under exceptional circumstances and depending on the concrete modalities of the processing." Grenzen ergeben sich dann nur noch aus dem Gebot der Zweckbindung (S. 3), sowie prozedural in Form von "adequate safeguards, such as providing individuals of electronic communication services the right to a judicial remedy" (ebenda, S. 2).

111 Zum grundrechtlich bes. wichtigen Kriterium des Anlasses siehe die *Kennzeichenkontrollen-2*-Entscheiung in BVerfGE 15, 244 Rn. 51, 93f.

112 Dieses Bundesjustizministerin *Christine Lambrecht* zugeschriebene Einschätzung zu verpflichtenden Tracking-Methoden wird zitiert bei *Grunert*, F.A.Z. v. 28.3.2020, S. 6.

113 So ausdr. *Kugelmann*, F.A.Z. v. 9.4.2020, abrufbar unter https://www.faz.net/aktuell/politik/staat-und-recht/was-juristen-und-datenschuetzer-ueber-die-corona-app-sagen-16718015.html <24.4.2020>.

chen Missbrauchs in Abwägung mit *konkreten* staatlich auferlegten Belastungen weiterhin ein bestimmendes Gewicht zugewiesen werden darf,[114] oder ob das Datenschutzrecht seinen eigentlichen Zweck damit nicht verfehlt, der nicht sein kann, die Vertrauenswürdigkeit demokratischer Institutionen einfach zu negieren, sondern dessen Aufgabe es sein sollte, diese Vertrauenswürdigkeit zu wahren.

3. Eigenverantwortung. Oder: disruptives Recht i. e. S.

Der dritte Punkt lässt sich vielleicht am besten mit der sehr abstrakten Frage einleiten: Was „soll" künstliche Intelligenz eigentlich?

Nach meinem Verständnis, und wiederum auf hoher Abstraktionsebene gesprochen, soll KI vor allem eines: Sie soll Menschen Verantwortung abnehmen; *Verantwortung* hier sehr weit verstanden als *Entscheidungs- und Handlungslast*. KI verspricht in diesem Sinne ein Mehr an Bequemlichkeit, etwa, wenn der digitale *personal assistant* uns nicht nur den schnellsten Reiseweg in den Urlaub vorschlägt, sondern dazu auch das am „besten" zu uns passende Hotel (nach welchem Bewertungsmaßstab auch immer); oder, wenn dieser *assistant* aus unseren Bank- und sonstigen Finanzdaten uns vollautomatisiert die Einkommenssteuererklärung erstellt. Alternativ soll KI ein Mehr an Lebensqualität schaffen, indem eine Medizin-KI personalisierte Therapien vorschlägt, die genauer zu unserer Erkrankung passen, Nebenwirkungen reduzieren usw.;[115] oder KI verspricht ein Mehr an Sicherheit und Regelbefolgung, indem sie Grenzbeamten hilft, genau die

---

114 Manche DiskussionsteilnehmerInnen scheinen das Mittel realer (physischer) Freiheitsbeschränkungen den mit Datenverarbeitungen potentiell (nämlich im Fall staatlichen Missbrauchs nach dem Vorbild einiger asiatischer Staaten) verbundenen Belastungen tatsächlich vorzuziehen. Es ist bedenkenswert zu fragen, ob hinter dem damit krass teuer (rechtlich und wirtschaftlich) erkauften Widerstand, von diesen asiatischen Staaten bewusst *selektiv* zu lernen, eine Form „zivilisatorischer Kränkung" stecken könnte. Mit dieser These *Siemons*, F.A.S. v. 29.3.2020, S. 41. Dahinter könnte freilich auch stehen, dass die verordneten Kontaktsperren zwar *rechtlich* real und auf physische Wirksamkeit angelegt waren, die *faktische* Möglichkeit zur Devianz aber sehr weitgehend unberührt gelassen haben; bei *tracking*-Apps wäre das vermutlich anders, die *tatsächliche* Freiheit wäre der rechtlichen Freiheit angepasst. Gedanken dazu auch bei *Rademacher* (Fn. 6).

115 Zu medizinischen Einsatzfeldern von KI siehe, jeweils mit m. w. N., *Jabri*, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 307ff., sowie *Molnár-Gábor*, ebendort, S. 337ff.

Personen herauszuwinken, die tatsächlich Drogen schmuggeln;[116] oder indem der digitale Steuerberater dafür sorgt, dass wir nicht – natürlich nur leichtfertig – ein Arbeitszimmer an- und absetzen, obwohl wir eine Einraumwohnung haben.

Unbeschadet des frühen technologischen Entwicklungsstadiums, will heißen: unbeschadet des Bewusstseins, dass viele der soeben angedeuteten Technologien (noch) nicht gut funktionieren, scheint es mir gewinnbringend, um des Arguments willen anzunehmen, *dass* die eben genannten Technologien perspektivisch gut funktionieren werden. Dass „die KI" also verlässlich und menschlich nachvollziehbar arbeitet, und auch für strafrechtliche Zwecke noch hinreichend klar wäre, welcher Mensch wie und wo intervenieren müsste und könnte, um wahrgenommene Fehlleistungen der KI zu verhindern oder zumindest rasch zu korrigieren.[117] Wäre die Architektur dann fertig, vollständig?[118]

Hier sind Zweifel angezeigt: Denn diese „schöne neue KI-Welt" würde immer noch drei Fragen, oder vielleicht besser formuliert: drei zentrale Wünsche offenlassen bzw. überhaupt erst aufwerfen:

- *Erstens*: Wie schaffen wir es, dass der Mensch analoge Fähigkeiten behält, z. B. in der medizinischen Diagnostik? Ich setze voraus, dass wir eine solche Fähigkeitserhaltungs-Architektur wollen.[119] Frei nach *Marietta Auer* zitiert sollte tatsächlich überlegt werden, ob es Bereiche

116 Hier sei etwa auf das Projekt *iBorderCtrl* verwiesen, dazu in diesem Band → *Wischmeyer*, bei S. 75. Fn. 15.

117 Natürlich sind wir davon derzeit noch weit entfernt. Allerdings wäre es verfehlt, die Diskussion auf die – derzeit noch deutlich sichtbaren – Defizite von KI (*false positives*, *black box*-Phänomen, Diskriminierungspotential etc.) zu beschränken; „[d]enn je mehr man strafrechtstheoretisch die Blößen von KI betont, desto mehr Anreize setzt man, dass solche Schwachstellen durch technologischen Fortschritt geschlossen werden." Dieses strafrechtswissenschaftliche Zitat von *Burchard* (Fn. 84), S. 19, dürfte *mutatis mutandis* für alle Rechtsdisziplinen gelten.

118 Das literarische Vorbild einer derartigen KI unter demokratisch-grundrechtlichen Prämissen wäre dann nicht mehr *George Orwells* „1984", sondern eher die paternalistisch-umsorgende KI aus *Bijan Moinis* „Der Würfel", 2. Aufl. 2019.

119 *Herrmann/Stock*, Kompetenzverlust in Zeiten von KI: Wie bewahren wir Menschen wichtige Fähigkeiten, in Ausgabe 2/2020 der Schriftenreihe *#Verantwortung KI – Künstliche Intelligenz und Gesellschaftliche Folgen* der IAG Verantwortung an der Berlin-Brandenburgischen Akademie der Wissenschaften, S. 24ff. Vgl. auch schon *Mayer-Schönberger/Ramge,* Das Digital, 2017, S. 260: „Daraus entstehen eine Reihe neuer Fragen, oft mit ethischer Dimension: Welche Entscheidungen wollen wir in jedem Fall selbst treffen, und welche können wir problemlos delegieren?".

gibt, wo wir als Gesellschaft in der Lage sein wollen, wieder „den Stecker ziehen“[120] zu können.[121]

- *Zweitens*: Wie schaffen wir es, dass der Mensch nicht auf eigenen/selbst- und biologisch festgelegten Pfaden „festhängt“, z. B. bei der Partnersuche mittels digitaler Apps, beim Medienkonsum[122] oder bei der Studienwahl anhand digitaler Studieneignungstests?
- *Zuletzt*: Wie schaffen wir es, dass der Mensch demokratisches Herrschaftssubjekt bleibt und sich als Herrschaftssubjekt wahrnimmt, statt von den ihn umgebenden digitalen Infrastrukturen physisch gesteuert zu werden (das Stichwort hierzu lautet *impossibility structures*)?[123] Sollen wir zum Beispiel, *in Maßen*,[124] die Freiheit zu zivilem Ungehorsam erhalten,[125] oder generell zu rechtlich deviantem Verhalten?[126] Rechtsverstöße müssen dafür in einer digitalisierten Zukunft vielleicht tatsächlich, *in Maßen*, als „Kritik am *status quo* anerkannt“[127] und damit ihre

120 *Auer*, Interview mit *Maximilian Steinbeis* v. 29.1.2020, abrufbar unter https://verfassungsblog.de/was-mich-eigentlich-interessiert-ist-das-gesellschaftliche/ <12.2.2020>.

121 Vgl. auch *Gaede* (Fn. 40), S. 76ff., mit der plausiblen Forderung, jedenfalls für die Rechtsdurchsetzung *gegen starke* intelligente Agenten analoge Reservekapazitäten vorzuhalten.

122 Die Frage hat auch überindividuell Bedeutung für den demokratischen Prozess, siehe *Sunstein*, #Republic, 2017, passim; *Schemmel*, Der Staat 57 (2018), 501ff.

123 Dazu m. w. N. *Rademacher* (Fn. 6), S. 707ff., und *ders*. (Fn. 50), S. 397f. Aus der strafrechtlichen Perspektive zudem *Fateh-Moghadam* (Fn. 3), S. 870ff., und – krit. – *Rostalski*, GA 2019, 481ff., bes. 485, die freiheitsphilosophisch argumentierend die Freiheit zum Bekenntnis *„für* das Recht“ (womit als andere Seite derselben Medaille freilich die Freiheit zum Handeln *gegen* das Recht verbunden ist), für eine rechtsstaatlich-freiheits(grund)rechtliche Fundierung der Diskussion plädiert. Für eine rechtsstaatliche Fundierung der Möglichkeit der Devianz tendenziell auch *Möllers* (Fn. 126), S. 478. Das erscheint mir zweifelhaft, jedenfalls wenn man rechtsstaatlich-dogmatisch argumentierend rechtliche Freiheit als Freiheit zum Handeln *innerhalb* von Rechtsbindungen konzipiert; hier ist es daher m. E. überlegenswert, ob die faktische Freiheit zum Handeln *außerhalb* von Rechtsbindungen nicht *demokratie*-, also herrschaftsrechtlich rekonstruiert werden müsste, dazu *Rademacher*, Vom rechtlichen Wert des Zufalls in der KI-Gesellschaft, Recht im Kontext-Vortrag an der Humboldt-Universität zu Berlin, 18.11.2019, S. 25ff., Manuskript beim Verfasser erhältlich.

124 Die Schwierigkeiten dürften darin bestehen, dieses Maß festzulegen, dazu *Rademacher* (Fn. 6), S. 709f.

125 *Becker*, ZUM 2019, 636.

126 *Möllers,* Die Möglichkeit der Normen, 2018, S. 476ff.

127 Zitat sowie sehr überzeugende Fragen hierzu bei *Burchard* (Fn. 84), S. 29 bzw. 24f.: „Die eigentliche Herausforderung für die Strafrechtstheorie liegt darin, dass

faktische Möglichkeit positiv konnotiert[128] oder gar verfassungsrechtlich geschützt[129] werden.[130]

Diese Fragen können in diesem Beitrag nicht vertieft werden, und es sei auch sofort eingeräumt, dass diese Fragen ganz unterschiedliche Rechts- und Schutzgüter adressieren[131] und vor allem nach Rechts- und Sachgebiet differenzierend feinjustiert werden müssen. Relativ naheliegend und schon jetzt diskutiert ist zum Beispiel ein Vorschlag für den Bereich der Medizin-KI, in digitale Diagnostiksysteme bewusst randomisierte Falsch-Diagnosen einzubauen, um den behandelnden Arzt weiter aufmerksam zu halten, ihm weiterhin ein Gefühl von Verantwortung zu erhalten.[132] In diesem und anderen Bereichen wäre zudem denkbar, Mensch und Maschine gezielt redundant und „im Wettbewerb" arbeiten zu lassen, wo dies ohne Effizienzverluste möglich ist.

Was den Antworten auf alle drei Fragen immerhin gemein sein könnte, ist Folgendes: Die neue KI-Verantwortungsarchitektur erfordert Bausteine, die die menschlichen Nutzerinnen und Nutzer – bei allem vielleicht ja einmal eingelösten Optimierungspotential der Technologie – in einem untechnischen Sinne *wach* und aufmerksam halten, sodass sie auf das Unerwartete, Unberechnete und Kontingente nicht nur reagieren können (wie

---

sich KI prima facie die zentralen Versprechungen des Strafrechts zu eigen macht – und sie optimiert. […] Eine grundsätzliche strafrechtstheoretische Kritik am Einzug von KI in die Strafrechtspflege müsste also nicht nur deren ‚fremde', sondern vielmehr die ‚eigenen', von KI ja vermeintlich nur übernommene Zielsetzungen kritisch in den Blick nehmen."

128 I. S. eines rechtspolitischen Petitums, die Rechtsvollzugsphase als Ort einer – dann v. a. vor Gericht und nicht qua Selbsthilfe auszutragenden – Kommunikation über das im konkreten Fall „richtige" Recht zu erhalten.

129 I. S. eines rechtsstaatlich und/oder demokratisch begründeten Gebots, dem Einzelnen eine Mitentscheidungsmöglichkeit über die Beachtung des Rechts zu geben (siehe die Nachw. hierzu in Fn. 123). Zu auf den ersten Blick parallel laufenden Überlegungen gegen „digitale Eigenmacht" im Internet der Dinge durch Anwendung des possessorischen Besitzschutzrechts auf vernetzte Geräte siehe *Kuschel*, AcP 220 (2020), 98, bes. S. 109ff., 116ff.; ob hier auch die Wertungen von Zivil- und öffentlichem Recht parallel laufen, wäre näher zu untersuchen.

130 Es entbehrt nicht einer gewissen Ironie, dass gerade das Datenschutzrecht, dem es ja gerade um die Sicherung von Selbstbestimmtheit und Autonomie geht, mit Art. 25 DSGVO und dem darin normierten *privacy-by-desing*-Ansatz die weitgehende Einführung von *impossibility structures* fördert und fordert; vgl. hierzu auch *Cornelius* (Fn. 9), S. 57ff.

131 Für im Ansatz ähnliche Überlegungen siehe auch *Danaher* (Fn. 20), S. 112ff.

132 Natürlich muss gewährleistet sein, dass das System den Fehler aufdeckt, bevor ein Arzt/eine Ärztin danach handelt.

im Medizindiagnostikbeispiel), sondern dem Unerwarteten einen positiven, eben *weil* Eigenverantwortung erhaltenden Mehrwert zuschreiben.[133] Denn andernfalls droht eine immer weiter getriebene Optimierung einer technologischen Verantwortung*sentlastung* des Menschen in seine Verantwortungs*unfähigkeit* und *-unmündigkeit* abzugleiten. Anders gewendet: Es gilt, der Optimierung dort Grenzen zu setzen, wo ein Umschlagen vom *gut genug* in ein *zu gut* droht.[134] Man könnte dann von disruptiven Bausteinen oder disruptivem Recht im engeren Sinne sprechen. Bewusste Zufallsarrangements, in je nach Sachbereich unterschiedlich gewählten „Einstellungen", könnten damit in der Zukunft – nach der überkommenen Dogmatik[135] durchaus überraschende – rechtliche Legitimität erlangen.[136]

III. Fazit

Dieser Beitrag schließt mit einem sehr offenen Fazit in Form eines Appells: Die mit KI und Verantwortung befasste Rechtswissenschaft sollte perspektivisch ihre Überlegungen *mindestens auch* daran ausrichten, dass die *technologisch* begründeten Verantwortungsprobleme bzw. -herausforderungen rund um KI gelöst werden könn(t)en; jedenfalls wird von den technologisch-ingenieurswissenschaftlichen Disziplinen hart an Lösungen gearbeitet.[137] Viele der dafür notwendigen Bausteine sind bekannt, erkannt und werden nun auch langsam zusammengesetzt (→ I.2). Eine Maschinenverantwortlichkeit in einem anspruchsvollen, oben referierten Sinn (→ II.1.) brauchen wir dafür im Zweifel nicht; eine Verantwortung der Bürgerinnen und Bürger und Unternehmen als – äußerst unvornehm ausgedrückt – Datenlieferanten werden wir hingegen wohl brauchen und auch bekommen (→ II.2). Bei alledem sollte darauf geachtet werden, dass wir bei aller KI-gestützten Optimierung unserer Gesellschaft – wozu mit hoher Sicherheit auch die für KI-Anwendungen geschaffenen Verantwortungsstruktu-

---

133 Vgl. literarisch *Hermann*, in: Fecher (Hrsg.), Twentyforty – Utopias for a Digital Society, 2020, S. 211ff.

134 Mit Blick auf die Funktionslogik des Strafrechts in diesem Sinne schon *Burchard*, zitiert in Fn. 127.

135 Zum Zufall als Rechtsprinzip immer noch aktuell *Depenheuer*, JZ 1993, 171ff.

136 So auch *Möllers* (Fn. 126), S. 478; *Rademacher*, Vom rechtlichen Wert des Zufalls in der KI-Gesellschaft (Fn. 123).

137 Treffend *Burchard* (Fn. 84), S. 30: „[I]nterne informationstechnologische Zweifel an der Leistungsfähigkeit von KI [befördern] nur weitere Entwicklungsspiralen […]." Siehe im Einzelnen zu technologischen Fortschritten mit spezifischem Verantwortungsbezug oben, bei Fn. 75.

ren zählen werden[138] –, dass wir also bei aller technischen Verantwortungs*entlastung*, auch Bausteine verbauen, die unsere rein analogen, menschlichen Fähigkeiten zu eigenverantwortlichem Handeln fordern und damit erhalten (→ II.3).

138 Dazu oben, bei Fn. 73ff.