

## Die Abwägungsresistenz der Menschenwürde als Opfer der künstlichen Intelligenz

Roberto Simanowski\*

### I. Einleitung

Bevor sich ein Medienwissenschaftler zu einem juristischen Problemfeld äußert, empfiehlt sich der Hinweis, dass es oft die Medienwissenschaft ist, die sich von der Rechtswissenschaft eine Klärung ihrer eigenen Forschungsfragen verspricht. Denn während die Medienwissenschaft die definitorische Ungewissheit ihres Gegenstandes als ein produktives Defizit ansieht, an dem man durchaus festhalten sollte, muss die Rechtswissenschaft zumindest in ihrem Zweig der Rechtsprechung Definitionen generieren, die eine juristische Einordnung der neuen Medien zur Klärung von Streitfällen ermöglichen.

Zu solchen Streitfällen gehörte im Jahr 2018 die Verantwortlichkeit von Facebook, YouTube, Twitter und anderen Social Media Plattformen für Hassbotschaften, die über ihre Kanäle verbreitet werden, die Frage, ob Google und vergleichbare Plattformen für die Übernahme urheberrechtlich geschützter Inhalte Nutzungsgebühren zahlen müssen, und die Frage, ob Eltern Zugriffsrechte auf die Facebookseite ihrer Kinder haben, wenn diese plötzlich sterben und Facebook ihre Seite sperrt. Dieser letzte Fall, der erst in dritter Instanz im Interesse der Eltern entschieden wurde,<sup>1</sup> offenbart die Schwierigkeiten, in diesem durch die neuen Medien geschaffenen juristischen Neuland zu strapazierfähigen Urteilen zu kommen. Denn das Berliner Kammergericht hatte das ursprüngliche Urteil des Landgerichts, dass Facebook den Eltern Zugang zur Seite ihrer Tochter

---

\* promovierter Literatur- und habilitierter Medienwissenschaftler. Er lebt, nach Professuren an der Brown University, der City University of Hongkong und der Universität Basel, in Rio de Janeiro und Berlin. Derzeit ist er Distinguished Fellow of Global Literary Studies im Excellence-Cluster "Temporal Communities" an der Freien Universität Berlin. Er ist Autor u.a. von *Data Love* (2014), *Facebook-Gesellschaft* (2016) und *Todesalgorithmus. Das Dilemma der künstlichen Intelligenz* (2020), das die hier unternommenen Überlegungen ausbaut und 2020 den Tractatus-Preis für philosophische Essayistik 2020 erhielt.

1 BGH, Urt. v. 12. Juli 2018, Az. III ZR 183/17.

gewähren muss, mit dem Verweis auf die Geltung des Fernmeldegeheimnis aufgehoben.<sup>2</sup> Diese Argumentation ist doppelt paradox, weil die Logik des Fernmeldegesetzes nicht nur am falschen Medium eingesetzt wird, sondern auch im falschen Modus. Zum einen ist in sozialen Netzwerken eine Meldung (bzw. ein „Update“ oder „Posting“) sowohl der Adressatin als auch all jenen, denen diese als ihren „Freunden“ Zugang zu ihrer Facebookseite gewährt, zugänglich, und zwar einschließlich derer, die diesen Zugang erst nach Versendung der Meldung erhalten. Zum anderen gilt das Briefgeheimnis der Post nur bis zur Auslieferung des Briefes, erstreckt sich aber nicht auf die geöffneten Briefe, die dann in der Schatulle im Schrank des Empfängers liegen. Wenn, wie im vorliegenden Falle, die Eltern der Verstorbenen über das Passwort zu deren Facebook-Account verfügen, ist das vergleichbar dem Umstand, dass sie Zugang zum Schrank und zur Schatulle im Zimmer der toten Tochter haben. Wenn Facebook dann den Account in den Gedächtnismodus einfriert und den Zugang der Eltern blockiert, kommt dies der Absurdität gleich, dass die Post die Brief-Schatulle aus dem Zimmer ihrer verstorbenen Kundin holt.

Schon diese Streitfälle von völliger neuer Dimension im Bereich des Erb-, Urheber- und Kommunikationsrechts bezeugen, dass Technologien die Gesellschaft transformieren, womit, in der Konsequenz, auch das Recht im Wandel ist. Die längst bekannten Stichworte für diese Transformation, mit denen sich auch die Rechtswissenschaft auseinandersetzen muss, lauten Hacking, Cybermobbing und Identitätsdiebstahl. Zu den Stichworten der Zukunft gehören predictive policing, facial recognition, data mining und ransomware. Nachdem die Weltgesundheitsorganisation im Sommer 2018 die Computerspielsucht in die Liste seiner „mental health conditions“ aufnahm, ist zu vermuten, dass sich die Rechtswissenschaft im Kontext des Suchmittelgesetzes bald auch mit dem Like-Button und anderen technischen Features des Verführungsdesigns zur Erzeugung von Dopamin-Feedback-Schleifen in sozialen Netzwerken beschäftigen wird.

All diese Stichworte stehen dafür, dass neuen Technologien neue Handlungsmöglichkeiten schaffen, deren Rechtsrahmen zu klären ist. Dabei geht es nicht nur um neue Antworten auf alte Fragen. Es geht auch um ganz neue Fragen. So lange zum Beispiel keine Zukunftsaussagen möglich waren, bestand auch nicht die Frage, ob jemanden, der ein Verbrechen erst noch begehen wird, schon in Gewahrsam genommen werden darf. So lange man nicht mit einer App vom Äußeren eines Menschen auf dessen Cha-

---

2 KG Berlin, Urt. v. 31. Mai 2017, Az. 21 U 9/16.

rakter schließen konnte, verletzte das genaue Anschauen eines Menschen auch nicht dessen Persönlichkeitsrechte. Ebenso stellt sich die Frage, ob man ein zahlungsrelevantes Urheberrecht an den Daten hat, die man Tag für Tag allein dadurch produziert, dass man lebt, erst dadurch, dass diese Daten nun in sozialen Netzwerken und im Internet der Dinge massenhaft abgegriffen und per Algorithmen in markttaugliche Serviceangebote verwandelt werden können. Und der Netzwerkeffekt verlangt einen völlig neuen Ansatz im Kartellrecht, denn er produziert zwar wettbewerbsfeindliche Datenmonopole, die aber nicht zu Preissteigerungen führen, sondern den ohnehin kostenlosen (und indirekt finanzierten) Service verbessern.

Wenn wir sagen, neue Technologien schaffen neue Handlungsmöglichkeiten und erfordern entsprechend neue Rechtsnormen, ist in medien- und sozialwissenschaftlicher Perspektive ein entscheidender Zusatz unerlässlich: Diese neuen Handlungsmöglichkeiten sind keineswegs alle gewollt oder vorhersehbar. Die Soziologie weiß dies nicht erst seit Robert K. Mertons Aufsatz „The Unanticipated Consequences of Purposive Social Action“ aus dem Jahr 1936, der auf die Probleme Wissensmangel, Fehleinschätzung und Kurzsichtigkeit verweist, wenn es um die Folgenabschätzung sozialer Handlungen geht. Schon Georg Simmels Essay „Der Begriff und die Tragödie der Kultur“ aus dem Jahr 1911 betonte, dass der Mensch sich mit seinen Erfindungen Sachordnungen schafft, die ihm – und zumal den künftigen Generationen – als äußere Kräfte, denen er sich nicht entziehen kann, gegenüberreten. So wird der Mensch, das ist die Tragödie, die Simmel in der Kulturentwicklung sieht, mit zunehmender Welteroberung zunehmend Opfer seiner Macht. Es ist dies eine Tragödie, nicht nur, weil künftige Generation keine Rechtsgrundlage haben, ihre Vorfahren für die geschaffenen Sachzwänge zu verklagen, sondern auch, weil der Mensch trotz dieser düsteren Aussicht nicht anders kann als im Schaffen neuer Technologien – und Sachordnungen – fortzufahren.

Wir werden auf den geschichtsphilosophischen Aspekt dieser Spannung zwischen Kultur und Technik zurückkommen. Schon hier aber sei festgehalten, dass solche Einsichten der Soziologie, die sich ebenso in der Medienwissenschaft manifestieren, für die Rechtswissenschaft insofern von Belang sind, als sie mit Blick auf rechtsrelevante Folgen technischer Entwicklungen die Möglichkeit der Verantwortungszuschreibung prinzipiell in Frage stellen. Es stimmt zwar, dass sich im Code auch Kultur äußert, und insofern sind Programmierer für die Wenn-Dann-Logik, die sie den Algorithmen einschreiben, in gewissem Maße verantwortlich, einschließlich der Diskriminierungen, die daraus entstehen mögen. Aber auch in diesem Falle müssen „unanticipated consequences“ in Rechnung gestellt

werden, inklusive Diskriminierungen, die weder intendiert waren noch gewollt sind.

Ein Beispiel der jüngsten Geschichte ist der Algorithmus, den Amazon zur automatischen Bewertung von Bewerbungen einsetzte, um diese vorab mit eins bis fünf Sternen zu versehen. Wie sich erwies, bevorzugte der Algorithmus männliche Bewerber, weil ihm zum Training Bewerbungsschreiben der letzten zehn Jahre eingegeben worden waren, die vorrangig von männlichen Bewerbern stammten und ein Vokabular enthielten, das von Frauen weniger verwendet wird.<sup>3</sup> Dieses Beispiel ist nicht so alarmierend wie es scheint, denn es bezeugt kein systemisches Problem des maschinellen Lernens und auch keine kulturellen Vorurteile der Programmierer, sondern nur eine erstaunliche Unprofessionalität im Umgang mit den verfügbaren Daten. Um so mehr aber verdeutlicht der Vorfall die möglichen menschlichen Folgekosten solcher Automatisierungen und die Aufgabe der Rechtspraxis zu klären, wen genau eine diskriminierte Bewerberin verklagen könnte.

Die Verantwortungsfrage für die Folgen technischer Entwicklungen weist allerdings noch eine weitere Stufe auf. Man muss unterscheiden zwischen der menschlichen Normsetzung *durch* Technik und den Normen, die *aus* der Technik selbst folgen: den Sachordnungen der Technik, wie der Soziologe Simmel es formulierte, den Postulaten, die die Medien ihren ahnungslosen Nutzern aufdrängen, wie später der Vater der Medienwissenschaft Marshall McLuhan sagen wird. Diese Postulate und Sachordnungen können, da sie dem technischen System und nicht dem Rechtssystem entstammen, zwar anklagefrei, aber nicht sanktionslos gebrochen werden. Wer etwa auf Facebook komplexe Einträge von epischer Länge postet oder Textmessages erst am nächsten Tag beantwortet, fällt unvermeidlich

---

3 J. *Dastin*, Amazon scraps secret AI recruiting tool that showed bias against women, Reuters vom 10. Oktober 2018, [www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G](http://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G) (alle Internetquellen zuletzt am 10. Dezember 2020 aufgerufen): "The group created 500 computer models focused on specific job functions and locations. They taught each to recognize some 50,000 terms that showed up on past candidates' resumes. The algorithms learned to assign little significance to skills that were common across IT applicants, such as the ability to write various computer codes, the people said. Instead, the technology favored candidates who described themselves using verbs more commonly found on male engineers' resumes, such as "executed" and "captured," one person said. Gender bias was not the only issue. Problems with the data that underpinned the models' judgments meant that unqualified candidates were often recommended for all manner of jobs, the people said."

durch die Kommunikationsmaschen der sozialen Netzwerke. Die ungeschriebenen Gesetze der Technik sind nicht weniger wirkungsvoll als die ungeschriebenen Gesetze der Gesellschaft.

Weniger harmlos als dieses Beispiel ist eine Software wie *Faception*, mit der ein 2014 gegründetes Start Up aus Tel Aviv verspricht, fremden Menschen auf einen Blick anzusehen, ob es sich um „a potential pedophile, an aggressive person, or a criminal“ handelt.<sup>4</sup> Die Argumentation basiert auf der Kombination zweier zentraler wissenschaftlicher Annahmen: „genetics played a large role in determining face shape“ und „genes play a greater role in determining key personality traits like social skills and learning ability than the way we are brought up by our parents“.<sup>5</sup> Die Produktwerbung wendet sich naturgemäß primär an Sicherheitsdienste, geht aber durchaus darüber hinaus: „Being able to utilize facial images to also answer the questions ‘What is this person? What are his personality traits? What motivates him?’ could revolutionize how companies, organizations and even robots understand people, and dramatically improve public safety, communications, decision-making, and experiences.“<sup>6</sup> Die implizierten Nutzungsformen jenseits öffentlicher Risikobereiche – die algorithmische Durchleuchtung von neuen Mitgliedern eines Fitnessstudios, potentiellen Partnern auf einer Dating-Plattform, Bewerbern auf eine Arbeitsstelle oder Wohnung – deuten auf die angesprochenen Sachordnungen, die sich jenseits der ursprünglichen Zwecksetzung durchsetzen, in Bereichen des gesellschaftlichen Lebens, wo das ethisch höchst problematische Screeningverfahren nicht mehr mit dem üblichen Freiheit-gegen-Sicherheit-Versprechen verteidigt werden kann. Lässt sich eine solche Zweckentfremdung gesetzlich verbieten? Lässt sich der Einsatz oder schon die Herstellung bestimmter Software juristisch regulieren? Muss hier die Legislative tätig werden, um der Judikative entsprechende Handlungsmöglichkeiten zu schaffen?

Auch das Beispiel *Faception* wirkt allerdings harmlos angesichts der Programmierung von Algorithmen, die sich schließlich ihre eigenen Gesetze schreiben beziehungsweise im Verfahren des „deep learning“ zu ihren eigenen Schlüssen kommen. Was soll man von einer Erfindung halten, die möglicherweise die letzte des Menschen sein wird, weil alles künftigen Erfinden – und die damit einhergehenden Sachordnungen – von ihr übernommen wird?

---

4 *Faception*, Verticals, <https://www.faception.com/hls-and-public-safety>.

5 *Faception*, About Us, <https://www.faception.com/about-us>.

6 *Faception*, About Us, <https://www.faception.com/about-us>.

Mit dieser Frage bin ich beim zentralen Thema meines Aufsatzes: künstliche Intelligenz (KI) bzw. artifizielle Intelligenz (AI). Ausgangspunkt meiner Überlegungen ist ein bestimmter Anwendungsbereich der KI, das selbstfahrende Auto, das von einem Algorithmus gesteuert wird, der im Kollisionsfall nach vorprogrammierten Regeln über Leben und Tod entscheidet. Insofern dieser Algorithmus mit einer spezifischen Entscheidungsmoral ausgestattet werden muss, wird die Philosophie zum wichtigsten Element in der Produktionskette eines Automobils. Allerdings ist das selbstfahrende Auto nur die Testfahrt für eine prinzipielle Transformation des moralisch-rechtlichen Selbstverständnisses der Gesellschaft. Der Algorithmus am Steuer, so die hier zugrunde liegende These, ist der Pilot der kommenden Gesellschaft, die eine kybernetische sein wird. Die ethischen Implikationen dieser angenommenen Entwicklung sind hier weniger aus einer rechtswissenschaftlichen als aus einer medienphilosophischen Perspektive zu beleuchten, bleiben aber einer rechtsphilosophischen Zuspitzung der Herausforderung des Rechts durch die Technik der Zukunft verbunden. Wir nähern uns dem Thema mit einer Fahrt in die Vergangenheit, in einem Mercedes der Luxusklasse.<sup>7</sup>

## II. Predictive Policing

In der wahrscheinlich berühmtesten Fake-Werbung fährt ein Auto durch eine Landschaft, die deutlich in der Vergangenheit liegt. Alle Menschen halten in der Arbeit inne und schauen mit Ehrfurcht auf etwas, das sich schließlich als ein Mercedes der Luxusklasse erweist. Kurz darauf stoppt dieser Mercedes vor zwei Mädchen, die auf der Straße spielen. Der Bremsvorgang geht offenbar auf einen Sensor zurück, der Objekte vor dem Fahrzeug wahrnimmt. Als das Auto weiterfährt, nun unterlegt mit spannungsgeladener Musik, kommt ein rennender Junge mit einem Drachen ins Bild und eine junge Frau, die ihm beim Wäscheaufhängen glücklich hinterherschaut. Das Zusammentreffen des Jungen mit dem Mercedes überrascht nicht, wohl aber der Zusammenprall. Hat diesmal der Sensor versagt?

Die Antwort kommt rasant in kleinen Stücken: Kurz vor dem Aufprall sieht man für eine Millisekunde das Bild von Hitler, dann ruft die Frau mit dem Wäschekorb erschrocken „Adolf!“, auf dem Ortseingangsschild

---

7 Die im Folgenden präsentierten Überlegungen basieren auf meinem Essay *R. Simanowski, Todesalgorithmus*, Wien 2020.

steht „Braunau am Inn“, aus der Vogelperspektive formen sich die Glieder des überfahrenen Jungen auf der Straße zu einem Hakenkreuz, der daraufhin eingeblendete Werbesatz für Mercedes' Bremssystem lautet: „Erkennt Gefahren, bevor sie entstehen“.<sup>8</sup>

Dieses Video aus dem Jahr 2013 hat im Frühjahr 2016 über 5 Millionen Views; mehr als 21 000 finden es gut, knapp 2 000 finden es schlecht, weil es die Tötung eines Kindes propagiert. Der Hinweis, dass es sich nicht um einen autorisierten Mercedes-Werbespot handelt, sondern um die Abschlussarbeit von Studenten der Filmakademie Ludwigsburg, verringert so wenig wie der Hinweis auf den völlig fiktionalen Rahmen das philosophische Problem, das hier aufgerufen wird: Unter welchen Umständen darf man töten, um Leben zu retten?

Die Fiktion spielt mit einem Topos der Popkultur, indem sie den Mercedes wie einst das Skynet den Terminator zurück in die Vergangenheit schickt, um den Beginn einer ungewollten geschichtlichen Entwicklung zu verhindern. Während Zeitreisen in die Vergangenheit allerdings reine Science Fiction sind, werden sie in die andere Richtung allmählich Teil unserer Gegenwart. Das Stichwort heißt vorhersagende Analyse, das in der Variante des „predictive policing“ an vielen Orten im täglichen Polizeibetrieb bereits zum Einsatz kommt. Anders als in Steven Spielbergs *Minority Report* basiert das Verfahren freilich nicht auf den hellseherischen Fähigkeiten dreier Frauen im Wasser, sondern auf Big Data Mining: auf der Erstellung von Tatprofilen, Korrelationen und Wahrscheinlichkeitskurven, die ‚berechnen‘ lassen, wann wo die nächste Straftat geschieht.

Werden solche statistischen Analysen mit den DNA-Werten, Video-, Lektüre- und Freundeslisten, Tagesroutinen und Bewegungsprofilen sowie sämtlichen Posts, Likes, Shares, Kommentaren und sonstigen kommunikativen Handlungen eines Individuums gekoppelt und entsprechend hochgerechnet, lässt sich bald vielleicht tatsächlich mit großer Genauigkeit die Entwicklung eines Menschen und damit sein gesellschaftliches Gefahrenpotenzial voraussagen. Mit solchen Daten versehen hätte der Mercedes die Wissenschaft auf seiner Seite, wenn er in Braunau auf den jungen Hitler stößt.

---

8 Im Herbst 2018 ist das Video unter der ursprünglichen YouTube-Adresse nicht länger auffindbar, wohl aber auf anderen YouTube-Sites als „Mercedes Benz ADOLF Spot“ oder „Mercedes spot 2013 (unofficial)“.

### III. Weichensteller-Dilemma

Die Frage, ob man töten darf, um Leben zu retten, ist längst in der Mitte der Gesellschaft angekommen, und zwar brutaler als im gezeigten Video, weil es nun gar nicht mehr um den Tot künftiger Mörder geht. In Ferdinand von Schirachs Gerichts-drama „Terror“ sitzt das Publikum über einen Major der Bundeswehr zu Gericht, der gegen die Weisung seiner Vorgesetzten ein von einem Terroristen entführtes Passagierflugzeug abschoss. Der Major entschied quantitativ und opferte die 164 Menschen im Flugzeug, um die 70 000 Menschen in der Münchner Allianz-Arena zu schützen, in die der Terrorist die Maschine jagen wollte. Die Entscheidung, ob der Major dafür wegen mehrfachen Mordes angeklagt werden soll, wird in den 2368 Aufführungen des Theaterstücks zwischen Oktober 2015 und Mai 2019 jeweils ans Publikum, als dem ‚realen‘ Schöffen in diesem Gedankenexperiment, delegiert.

Die Filmversion des Stückes war am 18. Oktober 2016 *das* Fernsehereignis in ARD, ORF und SRF: mit Publikumsabstimmung und anschließender Talk-Show *Hart aber fair*. Diese Verlagerung der Entscheidung ans Publikum spiegelt den Ansatz der experimentellen Ethik, die mit empirischen Studien operiert, statt sich, wie die sogenannte „Lehnstuhl-Philosophie“, mit theoretischen Schlussfolgerungen zu begnügen. Wenn solcherart Rechtsfindung auf Rechtsempfinden gründet, siegt nicht mehr das bessere Argument, sondern die Mehrzahl; zumindest lässt sich ohne diese im Rücken kaum noch argumentieren.<sup>9</sup> Das Ergebnis der Befragung zu von Schirachs Entscheidungsdilemma lautet jeweils Freispruch und zeigt: Die Mehrheit der Theaterbesucher (63 %) und der Fernsehzuschauer (87 %) denkt oder fühlt verfassungswidrig.

---

9 Vgl. die Dokumentation der Abstimmungsergebnisse auf *F. von Schirach*, Terror, <http://terror.kiepenheuer-medien.de>. Es erstaunt sicher nicht, dass die Rechtsgelehrten gegen ihre Entmachtung zugunsten einer juristisch völlig uninformierten Volksbefragung vehement protestieren, wie Strafrechtslehrer *W. Schild* in seiner Streitschrift: Verwirrende Rechtsbelehrung: zu Ferdinand von Schirachs „Terror“, Münster/Berlin 2016 und Bundesrichter *T. Fischer* in seiner ZEIT-Kolumne Schwarzer und Schirach, ZEIT Online vom 17. Mai 2016, [www.zeit.de/gesellschaft/zeitgeschehen/2016-05/literatur-strafrecht-fischer-im-recht](http://www.zeit.de/gesellschaft/zeitgeschehen/2016-05/literatur-strafrecht-fischer-im-recht) und „Terror“ – Ferdinand von Schirach auf allen Kanälen, ZEIT Online vom 18. Oktober 2016, [www.zeit.de/gesellschaft/zeitgeschehen/2016-10/ard-fernsehen-terror-ferdinand-von-schirach-fischer-im-recht/komplettansicht](http://www.zeit.de/gesellschaft/zeitgeschehen/2016-10/ard-fernsehen-terror-ferdinand-von-schirach-fischer-im-recht/komplettansicht). Zur experimentellen Ethik vgl. den Sammelband von C. Lutge/H. Rusch/M. Uhl (Hrsg.), *Experimental Ethics: Toward an Empirical Moral Philosophy*, New York 2014.

Denn es widerspricht dem Eingangsparagrafen der deutschen Verfassung über die Unantastbarkeit der Menschenwürde, Leben gegen Leben abzuwägen und gegebenenfalls wenige unschuldige Menschen zur Rettung vieler unschuldiger Menschen zu töten. Im Widerspruch zum moralischen Impuls der meisten Bürger besagt die ethische Grundlage der deutschen Rechtsprechung: Das kleinere von zwei Übeln lässt sich weder mathematisch ermitteln noch durch Diskriminierung nach Alter, Geschlecht oder kulturellen Wertmaßstäben. Das Leben von zehn Menschen ist nicht mehr wert als das von zwei und das Leben eines Kindes nicht mehr als das eines Greises. Die Ethik des unverhandelbaren Subjekts verbietet, einen Menschen auf ein Mittel zur Rettung anderer zu reduzieren.

Die deutsche Verfassung bevorzugt mit dem Verbot der Objektivierung oder Instrumentalisierung des Menschen die deontologische Pflichten- oder Gesinnungsethik gegenüber der konsequentialistischen Zweck- oder Verantwortungsethik. Während letztere auf das Ergebnis blickt und die Opferung der wenigen zur Rettung der vielen für vertretbar hält, akzentuiert erstere das Handeln (im vorliegenden Fall der Abschussentschluss des Majors) und bewertet die negative Pflicht, niemanden zu töten, höher als die positive Pflicht, Menschen zu retten. Aus genau diesem Grund kassierte das deutsche Verfassungsgericht Anfang 2006 Paragraph 14 Absatz 3 des vom Deutschen Bundestag ein Jahr zuvor verabschiedeten Luftsicherheitsgesetzes – der im Terrorfall als ultima ratio den Abschuss von Passagierflugzeugen erlaubte – als Verstoß gegen die Menschenwürde.

In der Philosophie wird das Dilemma des Tötens, um Leben zu retten, als „Weichenstellerfall“ diskutiert: Eine Straßenbahn ist außer Kontrolle geraten und droht fünf Personen zu überrollen; die einzige Handlungsoption eines Zeugen neben dem Weichenhebel besteht darin, die Bahn auf ein Nebengleis umzuleiten, wo sie nur *eine* Person überfahren würde. Der Tötungsbeschluss, den diese Weichenstellung enthält, wird in der sogenannten Fetter-Mann-Variante explizit, wenn man alternativ die Straßenbahn zum Halten bringen kann, indem man einen fetten Mann von der Brücke auf die Schienen stößt.

Philosophische Gedankenexperimente lassen sich auf keine Kompromisse ein. Deswegen fehlt im Weichensteller-Dilemma die Happyend-Variante der Selbstopferung. Man kann nicht an Stelle des fetten Mannes selbst auf die Schienen springen, um die Straßenbahn aufzuhalten. Im realen Leben ist das mitunter möglich, wenn man zum Beispiel in einem Unfall, vor die Wahl gestellt wird, das Auto gegen einen Erwachsenen, ein Kind oder eine Häuserwand zu fahren. Ob man zu dieser Entscheidung der Selbstopferung – die einzig moralisch und ethisch unbedenkliche Form der Tötung, um Leben zu retten – in der Lage ist, wird man erst wissen,

wenn es soweit ist. So jedenfalls verhielt es sich bisher. Anders wird es, wenn Algorithmen am Steuer sitzen.

#### IV. Instrumentalisierungsverbot

Der deutsche Verkehrsminister schuf im Mai 2016 eine Kommission unter Vorsitz eines früheren Bundesverfassungsrichters, um "ethische Fragen beim Paradigmenwechsel vom Autofahrer zum Autopilot" zu klären. Das im Juni 2017 vorgelegte Ergebnis untersagt für eine unausweichliche Unfallsituation strikt „jede Qualifizierung nach persönlichen Merkmalen (Alter, Geschlecht, körperliche oder geistige Konstitution)“ und verbietet eine „Aufrechnung von Opfern“, da das Individuum als „sakrosankt“ anzusehen ist.<sup>10</sup>

Im Gegensatz zur deutschen Perspektive, die konsequent am deontologischen Ethikmodell festhält, experimentiert in den USA das Massachusetts Institut of Technology in seiner *Moral Machine* – eine Plattform, auf der Besucher in dreizehn Dilemma-Szenarien eine Entscheidung treffen sollen – mit allen möglichen Klassifikationen. Hier muss man nicht nur wählen zwischen dem Tod eines Fahrzeuginsassen versus fünf Fußgänger oder drei Frauen und zwei Kinder versus fünf Rentner. Man soll auch entscheiden, ob der Tod von vier Insassen dem Tod von vier Passanten, die bei Rot die Straße überqueren, vorzuziehen ist oder der Tod von drei Kriminellen und einem Obdachlosen dem von zwei Doktoren und zwei Frauen. Damit geht die Verrechnungslogik über das Prinzip der Verletzungsminimierung hinaus und schließt den Verursachungsaspekt sowie den Reputationsindex mit ein.

Die *Moral Machine* versteht sich als „platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars“ und ist damit ein weiteres Beispiel für den Trend, ethische Fragen, die durch technologische Entwicklungen entstehen, empirisch-experimentell zu klären, also Rechtsfindung auf Rechtempfinden aufzubauen.<sup>11</sup> Abgesehen von dieser Verquickung von Ökonomie und Ethik besteht das methodische Problem des vorliegenden Ansatzes darin, diskussionslos den vor allem im englischsprachigen Raum verbreiteten konsequentialisti-

---

10 *Ethik-Kommission*, Automatisiertes und vernetztes Fahren, Bericht, Juni 2017, [www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?\\_\\_blob=publicationFile](http://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile), S. 11 und 18.

11 E. Awad et al., The Moral Machine Experiment, *Nature* 563 (2018), [www.nature.com/articles/s41586-018-0637-6#ref-CR9](http://www.nature.com/articles/s41586-018-0637-6#ref-CR9), S. 59–64.

schen Ansatz zu favorisieren. Denn in keinem der Szenarien gibt es die Möglichkeit der Stimmenthaltung: Eine Ablehnung der Zweckethik kann empirisch nicht erfasst werden. Dass die *Moral Machine* sich damit in einen Widerspruch zu den Prinzipien der deutschen Ethik-Kommission begibt, wird im Auswertungsbericht durchaus reflektiert und als unvermeidbar verteidigt. Wenn man der Wirtschaft Ratschläge geben wolle, wie sie die Algorithmen selbstfahrender Autos programmieren soll, könne man sich nicht der Entscheidung – oder eben „Aufrechnung von Opfern“, wie die Ethikkommission es nennt – entziehen.<sup>12</sup>

Dass der Rat auf Daten basiert, die quasi vor Ort erhoben wurden, ließe sich leicht verteidigen. Immerhin operiert der empirische Bottom-Up-Ansatz anders als das prinzipielle Top-down-Verfahren kontextsensitiv. Das Problem liegt allerdings tiefer. Es liegt im Pragmatismus, mit dem hier jene ethischen Prinzipien begünstigt werden, die der technische Fortschritt und das wirtschaftliche Interesse zu gebieten scheinen. Da der über Leben und Tod entscheidende Algorithmus nicht *nicht* programmiert werden kann, erweist sich die deontologische Ethik als unbrauchbar. Es sei denn, man entschärft sie durch die Adaption ihres Prinzips an die Erfordernisse der neuen Konstellation, etwa durch das Argument, dass die mögliche Tötung eines Menschen durch einen Algorithmus insofern keinen Verstoß gegen die Selbstzweckformel des deontologischen Ansatzes darstelle, als der Mensch, dessen Unfallrisiko durch die neue Technologie erheblich gemindert wird, der unmittelbare Zweck dieser Technologie bliebe: „Es läge dann kein grundsätzlicher Widerspruch zum kantischen Instrumentalisierungsverbot vor. Niemand könnte sich zum Mittel herabgesetzt fühlen, wenn seine Chancen, ein selbstbestimmtes Leben zu führen, durch autonome Autos mit einem Unfalloptimierungsalgorithmus verbessert würden.“<sup>13</sup> Die unvermeidliche Aufrechnung von potentiellen Opfern im Ernstfall wäre somit nur auf der Oberfläche des konkreten Falles ein Verstoß gegen das Instrumentalisierungsverbot; abstrahiert vom Einzelfall, handelte es sich hingegen um die logische Bedingung zur Aufrechterhaltung der deontologischen Zwecksetzung des Menschen.

---

12 E. Awad et al. (Fn. 11): „manufacturers and policymakers should be, if not responsive, at least cognizant of moral preferences in the countries in which they design artificial intelligence systems and policies. Whereas the ethical preferences of the public should not necessarily be the primary arbiter of ethical policy, the people’s willingness to buy autonomous vehicles and tolerate them on the roads will depend on the palatability of the ethical rules that are adopted.“

13 C. Misselhorn, Grundfragen der Maschinenethik, Stuttgart 2018, S. 192.

## V. Normierung durch Technik

Was die Hersteller der selbstfahrenden Autos betrifft, so entziehen sich diese noch der eigentlichen Diskussion und konzentrieren sich ganz auf technische Herausforderungen oder gesetzliche Rahmenbedingungen. Gelegentlich wird sogar der Diskussionsbedarf bezweifelt und die Fragestellung mit dem Argument abgewiesen: Wenn selbstfahrende Autos die Zeit für solche Entscheidungen haben, haben sie auch die Zeit, rechtzeitig zu bremsen, zumal sie ja streng zum risikoarmen Fahren programmiert werden. Wer mehr an die physische Schwerkraft als an die psychische glaubt, wird eine solche Lösung jedoch ausschließen und in der Kombination von Vernetzung und Geschwindigkeit *im* Auto sogar ein Dilemma orten, das viel komplexer ist als im Weichensteller-Modell: weil es (ähnlich der *Moral Machine*) dessen klinische Abstraktheit durch Kontextwissen ersetzt.

Gehen wir von folgendem Fall in einer nicht allzu fernen Zukunft aus: Ein Junge rennt mit einem Drachen auf die Straße, direkt vor einen autonom fahrenden Mercedes. Der Bordcomputer weiß durch Gesichtserkennung und Data Mining im Internet in Sekundenbruchteilen, dass die Ärzte der Frau am Straßenrand eine Lebenserwartung von zehn Monaten geben, dass der Fahrradfahrer neben der Frau zwei kleine Kinder hat, dass ein weiterer Passant, der ebenfalls als Kollisionsziel in Frage käme, das einzig verbliebene Kind seiner pflegebedürftigen Mutter ist und dass die Ausweichrichtung, in der sich kein Mensch befindet, über das gebrechliche Brückengeländer auf ein Bahngleis in 50 Meter Tiefe führt.

Welche Entscheidung auch immer der Computer trifft und welche Rolle auch immer dabei die Erkenntnisse aus dem Internet spielen mögen, die Verfassungswidrigkeit beginnt lange vor der Lebenswertanalyse im Gefahrenmoment. Die Algorithmen werden nach dem Befund der Sachlage ihre Reaktion an genau der Wenn-Dann-Logik ausrichten, mit der sie programmiert wurden. Manche Autos mögen also dem Kind, das den Unfall verursacht, nicht ausweichen, andere mögen zielsicher auf die Frau mit der geringen Lebenserwartung zusteuern, wieder andere aufs Brückengeländer.

Das moralische Dilemma ist eines der Reihenfolge: Weil die Entscheidung, die der Algorithmus eines selbstfahrenden Autos treffen wird, lange vor dem Antritt der Fahrt bestimmt wurde, also pro- statt reaktiv erfolgt, ist sie aus ethischer Perspektive immer schlechter als die eines Menschen, dessen reflexartige Reaktion ja keineswegs den Tatbestand einer bewussten Entscheidung im Sinne der Aufrechnung von Opfern erfüllt. Es hat seinen Grund, dass solche Fragen nach dem Verhalten im hypothetischen

Ernstfall nicht Teil der Führerscheinprüfung sind. Es kann kein richtiges Verhalten abgefragt werden, weil ein solches nicht festgelegt werden darf.

## *VI. Schuldig durch Technikgebrauch*

Die neuen Technologien bringen ethische, psychologische und politische Fragen mit sich, die noch gestern völlig unverständlich gewesen wären. Das betrifft die Frage der Vorbeugehaft, wenn per „predictive analytics“ eine Straftat aufgedeckt wird, die noch gar nicht begangen wurde. Das betrifft die Frage der Informationspflicht, wenn die DNA-Analyse eine unheilbare tödliche Krankheit voraussagt. Es betrifft die Entscheidung über Leben und Tod, die bisher dem situativen individuellen Reflex überlassen war und nun gewissermaßen vorprogrammiert werden muss. Wer wird die Todesalgorithmen in unseren Autos programmieren? Werden die Fahrzeughalter die Wahl haben? Wird sie beim Fahrzeughersteller liegen? Wird es verschiedene Algorithmen für verschiedene Fahrzeugklassen geben? Für verschiedene Automarken? Für verschiedene Länder? Wird die Politik der Wirtschaft die Entscheidung aus der Hand nehmen und (nach einer Volksabstimmung) einen bestimmten Algorithmus anordnen? Werden verschiedene Staaten verschiedene Algorithmen haben? Wird es einen UN-Beschluss geben – und einen Schwarzmarkt?

Es gehört zum Wesen von Technik, Standardisierungen zu schaffen, was sich schon im begrifflichen Ursprung ausdrückt: *téchne* als Methode. Technik im engeren Sinne eines Apparates oder technischen Systems überführt Handlungsweisen aus dem Bereich der individuell variablen Anwendung in den Modus des Expliziten und allgemein Verbindlichen. Genau dies geschieht, wenn Fahrverhalten und andere Prozesse, die bisher individuell und situationsbedingt geregelt wurden, automatisiert werden. Das damit einhergehende Problem der Normierung stellt sich in anderen Fällen weniger dramatisch, aber gleichfalls prinzipiell, wenn etwa dem automatisierten Staubsauger gesagt werden muss, ob ein Marienkäfer oder eine Spinne als Lebewesen verschont oder als Form von ‚Dreck‘ eingesaugt werden soll.

Wenn es um Menschenleben geht, ist das Prinzipielle ungleich dramatischer. Es liegt nahe, die ethische Ausstattung eines Autos nicht dessen Hersteller zu überlassen, der sich sonst mit moralisch bedenklichen Versprechen einen Wettbewerbsvorteil zu verschaffen suchen könnte, indem er (wie zum Beispiel Mercedes-Benz) seine Autos damit anpreist, dass sie im

Ernstfall immer der Rettung der Insassen den Vorrang geben.<sup>14</sup> Allerdings ist eine gesellschaftliche Regelung nicht weniger problematisch. Das zeigt die Warnung der erwähnten Ethikkommission vor einem „Paternalismus des Staates, bei dem eine ‚richtige‘ ethische Handlungsweise vorgegeben wird“.<sup>15</sup> Wie aber soll die Regelung erfolgen, wenn sie nicht mehr dem Individuum überlassen werden kann, wie es das Wertebild des Humanismus vorsieht? Die Antwort bleibt vorerst unklar. Empirisch belegt hingegen scheint, dass die meisten Befragten zwar dafür plädieren, im Ernstfall das Leben der Fahrzeuginsassen zu opfern, um das Leben anderer zu retten, selbst aber kein Auto kaufen würden, dem eine solche Opferethik einprogrammiert ist.<sup>16</sup>

So groß das ethische Dilemma der Todesalgorithmen auch ist: Ihnen *nicht* die Regie zu überlassen, wäre keine Lösung. Noch unmoralischer, als kühl und herzlos das Leben eines Kindes gegen das eines Rentners abzuwägen, wäre es, eine Technologie zu blockieren, die Zehntausende von Unfalltoten pro Jahr verhindert. So jedenfalls die Argumentation der Befürworter des autonomen Fahrens und die Rechnung des Utilitarismus, wonach alles moralisch ist, was Glück maximiert – oder eben Unglück minimiert. Die Statistik ist das unschlagbare Argument der neuen Technologien, ganz gleich welche ethischen Dilemmata sie mit sich bringen – selbst militärische Drohnen werden schließlich genau damit gerechtfertigt: weniger Kollateralschäden um den Preis ihrer konkreten Akzeptanz.

## VII. Transkulturelle Ethik

Genau hierin könnte das Argument dafür liegen, im Ernstfall immer die Fahrzeuginsassen zu opfern. Zwar trifft diese die geringste Schuld an einem Unfall, wenn sie nur über den Algorithmus vermittelt am Straßenverkehr teilnehmen. Zwar werden, insofern Algorithmen verlässlicher

---

14 Zur Aussage des Sicherheitsmanagers Christoph von Hugo über die Rettungs- und Opferungsstrategie von Mercedes-Benz vgl. *M. Taylor*, Mercedes autonomous cars will protect occupants before pedestrians, Auto Express, 2016, [www.autoexpress.co.uk/mercedes/97345/mercedes-autonomous-cars-will-protect-occupants-before-pedestrians?\\_ga=2.252642126.650054044.1549235659-1961590890.1549235659](http://www.autoexpress.co.uk/mercedes/97345/mercedes-autonomous-cars-will-protect-occupants-before-pedestrians?_ga=2.252642126.650054044.1549235659-1961590890.1549235659).

15 *Ethik-Kommission* (Fn. 10), S. 16.

16 *J. Bonnefon/A.Shariff/I.Rahwan*, Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?, 13. Oktober 2015, <http://arxiv.org/pdf/1510.03346v1.pdf> sowie *dies.*, The Social Dilemma of Autonomous Vehicles, *Science* 352 (2016), S. 1573–1576. Rahwan, Bonnefon und Shariff produzierten zugleich das Konzept für MITs Moral Machine.

operieren als Menschen, gerade jene die Unfälle verursachen, die nicht in einem Auto sitzen: das Kind, das dem Ball hinterher auf die Straße läuft, die abgelenkte Fußgängerin am Handy, der Fahrradfahrer, der das Gleichgewicht verliert. Aber es gilt zu unterscheiden zwischen der konkreten Schuld, die aus menschlichem Fehlverhalten resultiert, und jener generellen Schuld, die aus dem Technologiegebrauch folgt. Denn die eigentlichen Mobilitätsrisiken werden nicht von Fußgängern oder Fahrradfahrern erzeugt, sondern vom Autoverkehr, der schwere Gegenstände so schnell durch den Raum bewegt, dass Zusammenstöße lebensgefährlich werden. Auch wenn die Anzahl an Todesopfern durch selbstfahrende Autos gesenkt wird, das Todesopfer ist *an sich* eine Folge des Autoverkehrs. Verlangt die neue Technologie eine Vorentscheidung zur Ausweichstrategie des Fahrzeuges im Falle eines Unfalls, wäre es also durchaus angemessen, das Opfer den Nutznießern des Mobilitätsrisikos aufzubürden.

Eine andere Lösung wäre die Abstimmung: die Behandlung des ethischen Dilemmas mit quantitativen Mitteln, was die einen als Basisdemokratie befürworten, die anderen als Maximierungskalkül kritisieren würden. Denn schon die Abstimmung, ob das Prinzip der deontologischen Ethik zur Disposition stehen und im Ernstfall eine utilitaristische Aufrechnung von Menschenleben stattfinden sollte, würde – als quantitatives Verfahren der Entscheidungsfindung – dem Vermessungsprinzip des Utilitarismus den Vorrang geben. Gewinnen würde nicht das beste Argument, sondern die größte Zahl, das, was die größte Zustimmung erringt und damit, wenn nicht Glück, so doch Zufriedenheit mit dem Ergebnis maximiert.

Angesichts der nicht nur durch die *Moral Machine* bestätigten regionalen Unterschiede liegt es nahe, die Algorithmen unterschiedlich zu programmieren und gegebenenfalls die Abstimmung dazu regional vorzunehmen. In der Folge könnten, je nach moralischem Selbstverständnis und Mehrheitsentscheid, bestimmte Länder die Selbstopferung vorschreiben, andere streng utilitaristisch vorgehen und wieder andere der Rettung der Insassen den Vorrang geben, ganz gleich welche Gefahren dabei für andere entstehen. Natürlich könnte man die Algorithmen per GPS leicht auf die territorial jeweils geltenden Normen umkodieren, um die unterschiedlichen Wertvorstellungen national durchzusetzen. Aber soll VW fern ab vom deontologischen Deutschland in südlichen Ländern tatsächlich autonome Autos anbieten, die vor allem das Leben von „higher status charac-

ters“ und Haustieren schützen?<sup>17</sup> Hieße die regionale Entscheidung nicht, den moralischen Normen den Anspruch einer universalen Geltung und damit ihre moralische Rückversicherung zu entziehen?

Angemessener erscheint da der populäre Vorschlag, im Ernstfall einen Zufallsgenerator über die anzuwendende Norm entscheiden zu lassen. Auch das bliebe freilich unter der Erwartung, die sich an Automatisierungsprozesse koppelt. Denn Automatisierung heißt auch, das Besondere überwinden, sei es mit Blick auf Individuen oder mit Blick auf Kulturen. Die spannende Frage, die das Problemfeld des Todesalgorithmus enthält, ist also die nach der Möglichkeit, der global operierenden Technik einen ethischen Standard zu geben, der global verbindlich ist.

Bei einem solchen Anliegen geht es um mehr als die Beseitigung regionaler Differenzen. Ziel wäre auch die Versöhnung der gegensätzlichen Perspektiven einer Allgemeinen Ethik, die hier mit den Stichwörtern deontologisch und konsequentialistisch skizziert sind. Diese Versöhnung blieb in der Theoriebildung bisher aus, weil sich Gegensätze eben nicht vereinen lassen, wird auf der Ebene der angewandten Ethik nun aber durch das autonom fahrende Auto dringlich. Die Frage ist, ob und wie sich ein transkulturelles Einvernehmen erreichen lässt, ein Einvernehmen jenseits all der unterschiedlichen Wertvorstellungen, die bisher ja auch die Bemühungen um universell geltende Menschenrechte immer wieder behindern.<sup>18</sup> Bevor Algorithmen so programmiert werden können, dass sie gerecht operieren, ist freilich zu klären, was gerecht ist. Die verschiedenen Angebote müssen gegeneinander abgewogen und idealerweise handlungsfähig synthetisiert werden. Automatisierung verlangt Standardisierung. Das ist der Fall bei jeder halbwegs komplexen Maschine. Der Unterschied beim Todesalgorithmus: Hier handelt es sich nicht mehr um die Standardisierung des Handwerks, sondern des Denkens.

---

17 E. Awad et al., The Moral Machine Experiment, *Nature* 563 (2018), S. 59–64, [www.nature.com/articles/s41586-018-0637-6#ref-CR9](http://www.nature.com/articles/s41586-018-0637-6#ref-CR9).

18 Ein berühmtes Beispiel für die Unterschiede der Wertvorstellungen ist die Bangkok Declaration on Human Rights 1993, in der die „Asian Values“ gegen den „Menschenrechtsimperialismus“ des Westens in Stellung gebracht wurden. Werte und die daraus ableitbaren Rechte, das unterstrich diese Deklaration, sind nicht universell, sondern kulturell verschieden. Diese Verschiedenheit drückt sich schließlich auch in Umfragen zum Weichensteller-Dilemma aus, wonach wesentlich mehr US-Amerikaner als Chinesen bereit sind, aktiv in das Geschehen einzugreifen und die Weiche umzulegen. H. Ahlenius/T. Tännsjö, Chinese and Westerners Respond Differently to the Trolley Dilemmas, *Journal of Cognition and Culture* 12 (2012), S. 195–201.

### VIII. Geist des Werkzeugs

Die Lösung dieser Herausforderung liegt am ehesten im „deep learning“-Modell der Algorithmen. Diese Variante des maschinellen Lernens beruht auf einem selbstadaptiven Algorithmus, der sich von einem Instrument in den Händen der Menschen zu einem Akteur mit eigener Datenverarbeitungsstrategie und Entscheidungslogik emanzipiert. Ethische Kompetenz stellt natürlich vor andere Herausforderungen als die Fähigkeit der Spamerkennung, die der SpamAssassin entwickeln musste. Ein Ansatz, der am renommierten Georgia Institute of Technology verfolgt wird, besteht darin, dem KI-System literarische Texte, Filme und „crowdsourced material“ (via Amazons Mechanical Turk) vorzulegen, aus dem menschliche Entscheidungen in verschiedenen Verhaltenssituationen erkennbar werden. Auch diesem Ansatz liegt eine Logik der Quantität zugrunde: die Annahme, dass in der Masse die vorbildlichen Beispiele überwiegen und die künstliche Intelligenz somit die richtigen Schlüsse ziehen wird. Die Masse wiederum kann riesig sein, denn die Lektüre von Algorithmen erfolgt im Schnellverfahren des „distant reading“; der Lernprozess muss sich also nicht auf einen wie auch immer sortierten Kanon konzentrieren, sondern kann *alle* Objekte menschlicher Kultur einbeziehen.<sup>19</sup>

Ist das Informationsinput konsequent multikulturell organisiert und enthält, zum Beispiel, auch genügend Stimmen aus Afrika und von den Aborigines, müsste sich dies im Ergebnis niederschlagen: als Ausgleich der Differenzen oder als eine Art globaler Mehrheitsentscheid, der dann auch dem Todesalgorithmus universelle Gültigkeit verschaffen könnte. Der Nebeneffekt läge darin, dass das „Werkzeug des Geistes“ (so bezeichnete ein Technikphilosoph Anfang des 20. Jahrhunderts die Sprache) mit dem „Geist des Werkzeugs“, dem „Eigenwillen“ der Technik, außer Kraft gesetzt wäre.<sup>20</sup> Das menschliche Denken und Handeln würde wieder auf einen Nenner gebracht durch eine Technik, die selbst (als binärer Code) eine Sprache ist und einen eigenen ‚Geist‘ (den Geist des Rechnens) entwickelt.

Die Ethik des Todesalgorithmus, die sich aus diesem Geist im Idealfall schließlich jenseits einer ursprünglichen, auf bestimmten Werten basierenden Programmierung entwickelt, bedeutete im Grunde den Triumph der

---

19 S. Parkin, Teaching Robots Right From Wrong, *The Economist*, Juni/Juli 2017, S. 68–73; F. Moretti, *Distant Reading*, Konstanz 2016.

20 M. Eyth, Poesie und Technik, zit. n. E. Cassirer, Form und Technik, in: E. W. Ort/J. M. Krois (Hrsg.), *Symbol, Technik, Sprache: Aufsätze aus den Jahren 1927–1933*, Hamburg 1985, S. 39–91, hier: S. 74.

Technik über die Kultur in ihrer jeweiligen Ausprägung. Dieser Algorithmus führt geschichtsphilosophisch betrachtet zu einer Harmonie zurück, die ironischerweise gerade beim ersten großen Technikunternehmen der Menschen verlorenging: dem Turmbau zu Babel. Gott sah in diesem Turm, dessen Spitze „an den Himmel“ reichen sollte, eine Selbstüberhebung der Menschen: „Siehe, sie sind ein Volk und haben alle eine Sprache, und dies haben sie angefangen zu tun; und nun wird ihnen nichts verwehrt werden, was sie zu tun ersinnen.“ (Gen. 11,6) Als Gott zur Abwehr dieser Bedrohung die Sprache der Menschen verwirrte, mussten diese wegen der eintretenden Verständigungsschwierigkeiten ihr Bauvorhaben aufgeben und zerstreuten sich „über die ganze Erde“.

Insofern Sprache „Abdruck in der Seele [macht]“ und „das Haus des Seins [ist]“, wie Herder und Heidegger es formulierten, symbolisiert die babylonische Sprachverwirrung den Beginn kultureller Differenzierung, die – mittels Soziolekt, Ideologie, Diskurs und anderen Formen der Gruppensprache – durchaus auch *innerhalb* einer natürlichen Sprache entsteht.<sup>21</sup> Aus diesen Differenzen resultieren schließlich all die unterschiedlichen Realitätsauffassungen und Wertvorstellungen, die nicht zuletzt ein global operierender Todesalgorithmus miteinander zu versöhnen hätte. Denn anders als der Mensch ist der Algorithmus weder von kulturellen noch biologischen Faktoren abhängig, sofern er sich im Verfahren des „deep learning“ und „distant reading“ von seinen Programmierern emanzipiert. Tut er dies nicht, repräsentiert er deren Perspektiven und Vorurteile. In diesem Falle gilt der berühmte Tweet: „Machine learning is like money laundering for bias.“<sup>22</sup>

Die Möglichkeit einer solchen Emanzipation des Algorithmus von seinen Schöpfern wird im entsprechenden Fachdiskurs unterschiedlich bewertet. Gegen die Annahme, dass Algorithmen nie den Vorurteilen ihrer Programmierer entkommen können, stehen die Beispiele der ‚kulturellen Formung‘ von Chatbots wie etwa Microsofts *Tay*, das sich im März 2016 auf Twitter innerhalb eines Tages zu einem „Völkermord bejahende[n] Internet-Troll voller Hass und Hetze“ entwickelt hatte und damit natürlich nicht die Haltung Microsofts repräsentierte, sondern die politische Einstel-

---

21 J. G. Herder, Über die Wirkung der Dichtkunst auf die Sitten der Völker in alten und neuen Zeiten, in: Herders Sämtliche Werke, Bd. 8, Berlin 1969, S. 339; M. Heidegger, Holzwege, Gesamtausgabe Bd. 5, Frankfurt am Main 2003, S. 310.

22 Auf @Pinboard, 19. Juni 2016, <https://twitter.com/pinboard/status/744595961217835008?lang=en>.

lung der auf Twitter besuchten Nutzergruppen.<sup>23</sup> Gerade dieses Negativbeispiel bezeugt, dass künstliche Intelligenz auch jenseits der kulturellen Werte ihrer Programmierer ‚sozialisiert‘ wird. Diese Sozialisierung, so ist zu vermuten, erfolgt nicht nur körperlos, sondern auch umweltfrei insofern ihre Umwelt universell ist – was im Beispiel *Tay* gerade nicht der Fall war. Die Interaktionen der künstlichen Intelligenz mit potenziell allen Kommunikationsangeboten der Welt unterscheidet sich grundsätzlich von den sozialen Beziehungen menschlicher Intelligenz. Sie bleibt nicht an den Horizont einer bestimmten Sprache und an die Praxis bestimmter Diskurse gebunden, sondern inkludiert alle Kommunikationsangebote, denen sie ausgesetzt ist. Sie erzeugt kein religiös oder politisch vorgeformtes Bewusstsein und Selbstverständnis. Sie ist Technik jenseits der Kultur. Der tiefere Sinn der „Singularity“-These liegt in der technischen Überwindung kultureller Pluralität.

### *IX. Dilemma des Todesalgorithmus*

Die Ironie global operierender Algorithmen in autonomen Fahrzeugen läge nun allerdings darin, dass diese Technik trotz ihrer universellen Geltung zugleich jenen universellen Wert außer Kraft setzt, auf dem der Grundsatz der unantastbaren Menschenwürde beruht: die Gattungszugehörigkeit des Menschen. Die Menschenwürde ist ja gerade deswegen unantastbar, weil sie kein kulturelles, soziales, ja nicht einmal moralisches Kriterium in Rechnung stellt, sondern allein ein biologisches: die Tatsache, ein Mensch zu sein. Das bedeutet in der Konsequenz zum einen, dass auch jene *homo sapiens* moralisch und rechtlich schutzwürdig sind, die nicht über die Eigenschaft der Vernunft verfügen wie Embryonen oder dauerhaft komatöse Patienten. Zum anderen deckt dieser Schutz auch Kriminelle und Terroristen ab, denen man trotz ihrer Verbrechen biologisch das Menschsein nicht absprechen kann.

Konsequentialistisch gesehen lässt sich diese universelle Geltung der Unantastbarkeit im Sinne der Schutznormlehre mit dem Argument der Gefahrenbeseitigung und des Opferschutzes beziehungsweise erweiterter Notwehrrechte durchaus in Frage stellen und zum Beispiel die „Rettungsfolter“ rechtfertigen, um Terrorakte zu verhindern. Diese Haltung – die

---

23 A. Steiner, Was Microsoft durch „Tay“ gelernt hat, FAZ vom 26. März 2016, [www.faz.net/aktuell/wirtschaft/netzwirtschaft/was-microsoft-mit-dem-bot-tay-von-der-netzgemeinde-gelernt-hat-14146188.html](http://www.faz.net/aktuell/wirtschaft/netzwirtschaft/was-microsoft-mit-dem-bot-tay-von-der-netzgemeinde-gelernt-hat-14146188.html).

in den USA gängige Praxis ist und vom ehemaligen Vizepräsident Dick Cheney offensiv verteidigt wurde – ließe sich rechtstheoretisch zudem damit verteidigen, dass sich Terroristen im Grunde selbst außerhalb des Menschseins begeben haben und somit das Recht auf Unantastbarkeit ihrer Menschenwürde gegenüber der Schutzpflicht des Staates gegenüber seinen Bürgern verspielt haben.<sup>24</sup> Versteht man auf diese Weise das Menschsein nicht mehr biologisch, sondern normativ und erlaubt dem Staat, die Würde der Opfer gegen die des Täters auszuspielen,<sup>25</sup> nimmt man jedoch Abstufungen in der Unantastbarkeit der Menschenwürde vor, die sich letztlich kaum kontrollieren lassen. Die Argumentation, bis zu welchem Verbrechen- und Gefährdungsgrad ein Mensch noch die Unantastbarkeit seiner Würde verdient, wird je nach philosophischem, politischem, kulturellem Hintergrund unterschiedlich ausfallen, was die Idee des Universellen jenseits kultureller Vielfalt ad absurdum führt. Wie begegnet man der Ansicht, dass Sodomiten, Abtreibungsbefürworter oder Ungläubige sich außerhalb des Menschseins stellen, wenn dieses nicht mehr bedingungslos gilt?

Genau diese Relativierung der universellen Menschenwürde erfolgt allerdings durch einen Todesalgorithmus mit universeller Geltung. Denn der Algorithmus muss ja nach allgemeinen Kriterien handeln, die in jedem vergleichbaren Fall verlässlich zur gleichen Entscheidung führen, diese(s) Leben statt jene(s) zu retten beziehungsweise diese(s) für jene(s) zu opfern. Er muss nach allgemeinen Kriterien die gelegentliche Antastbarkeit der Menschenwürde klären, die Möglichkeit, diese/n Menschen zum Mittel der Rettung jenes/r Menschen zu degradieren. Die Degradie-

- 
- 24 *W. Brugger*, Vom unbedingten Verbot der Folter zum bedingten Recht auf Folter?, JZ 2000, S. 165–173, vgl. das Gedankenexperiment zum Opferschutz S. 168: „Hält ein Geiselnehmer die Pistole an die Schläfe der Geisel, darf er erschossen werden, wenn dies das einzig erfolgversprechende Mittel zur Rettung des Lebens der Geisel ist. Hat ein Geiselnehmer die tickende Bombe an der Geisel befestigt, und kann diese nur durch Anwendung von Zwang zur Preisgabe des Codes des Zündmechanismus gegen den Geiselnehmer gerettet werden, schließen die einschlägigen Rechtsnormen diese Maßnahme aus.“; *F. Wittreck*, Menschenwürde und Folterverbot: Zum Dogma von der ausnahmslosen Unabwägbarkeit des Art. 1 Abs. 1 GG, Die Öffentliche Verwaltung 65 (2003), S. 873–883; *J. Isensee*, Menschenwürde: Die säkulare Gesellschaft auf der Suche nach dem Absoluten, AöR 131 (2006), S. 173–218. Vgl. den Sammelband H. Brunkhorst (Hrsg.), Rückkehr der Folter: Der Rechtsstaat im Zwielficht?, München 2006. Zur staatlichen Legitimation der Folter in den USA vgl. *B. E. Harcourt*, Gegenrevolution: Der Kampf der Regierungen gegen die eigenen Bürger, Frankfurt am Main 2019/New York 2018.
- 25 *W. Brugger* (Fn. 24), S. 169.

rung auf der Grundlage eines quantitativen Vergleichs – so im Weichenstellerfall – bereitet letztlich – so die *Moral Machine* – die Relativierung von Lebenswert in qualitativer Hinsicht vor, zwischen Alten, Kindern, Obdachlosen, Kriminellen, „Unbescholtenen“ und „Ehrenhaften“.

Die Ersetzung der spontanen Entscheidung in einer konkreten Situation durch eine situationsunabhängige Regel – ein „allgemeines Gesetz“ im Sinne Kants – markiert die moralische Unmöglichkeit der KI in deontologischer Hinsicht. Um dem Todesalgorithmus überhaupt moralische Anweisungen, die als Grundlage einer allgemeinen Gesetzgebung taugen, geben zu können, wird man das Modell der deontologischen Ethik – ein Grundpfeiler der deutschen Leitkultur – aufgeben müssen. Man wird Modelle der „Abstufung im Unrecht“ entwickeln, die eine Pflicht zur Schadensminimierung postulieren und dazu die Tötung von Unschuldigen rechtfertigt.<sup>26</sup>

## X. Schlussbemerkungen

Solche Modelle sind das unscheinbarere, aber ebenso folgenschwere Pendant zur Aufhebung des Folterverbots. Denn auf dem Spiel steht mehr, als das hier gewählte Beispiel des autonomen Fahrens und die Referenz zum Weichensteller-Dilemma vermuten lassen mag. Auf dem Spiel steht das „Tabu“ der Unantastbarkeit der Menschenwürde.<sup>27</sup> Während die Infragestellung dieser Unantastbarkeit im Falle des aufgehobenen Folterverbots durch die weltpolitische Entwicklung mit ihren neuen Gefahren für das gesellschaftliche Zusammenleben motiviert ist, liegt der Grund im Falle des autonomen Fahrens in der technischen Entwicklung mit ihren neuen Konstellationen der Mensch-Maschine-Interaktion. So wie gesellschaft-

---

26 E. Hilgendorf, *Autonomes Fahren im Dilemma: Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätigen Kollisionsvermeidungssystemen*, in: E. Hilgendorf (Hrsg.), *Autonome Systeme und neue Mobilität: Ausgewählte Beiträge zur 3. und 4. Würzburger Tagung zum Technikrecht*, Baden-Baden 2017, S. 143–176.

27 J. Isensee (Fn. 24), S. 173–218. Dass die von Isensee und anderen attackierte Abwägungsresistenz der Menschenwürde nicht rhetorisch auf den Status eines „Tabus“ reduziert und als „Denkblockade“ denunziert werden darf, sondern aus den eigenen Ressourcen der Rechtswissenschaft argumentativ verteidigt werden muss und kann, begründen ausführlich und facettenreich M. Neumann/T. Gutmann/T. Weitin/B. Fateh-Moghadam, *Säkulare Tabus*, Berlin 2015. Angemerkt sei, dass die Menschenwürde im Sinne des kantischen Instrumentalisierungsverbots kein Faktor der Allgemeine Erklärung der Menschenrechte ist.

liche und technische Entwicklungen zuvor die Unantastbarkeit Gottes in Zweifel zogen, stellen sie nun die „Sakralisierung“ des Menschen zur Disposition.<sup>28</sup> Der „Kult des Individuums“ – Durkheims moralsoziologische Antwort auf den Verlust eines (religiösen) Kollektivbewusstseins im ausgehenden 19. Jahrhundert – wird relativiert durch die Logik der Zahl, die, wie die Reaktionen des Publikums auf das Gerichtsdrama *Terror* und auf das Experiment der *Moral Machine* zeigen, der Gruppe (an Geretteten) den Vorzug vor dem (unantastbaren) Individuum gibt.

Insofern die Unantastbarkeit der Menschenwürde darauf zielt, dem Einzelnen in seinem Eigenwert „einen nicht antastbaren Freiheitsbereich gegenüber Kollektivinteressen zu garantieren“,<sup>29</sup> bedeutet das mittlerweile staatlich praktizierte, rechtswissenschaftlich verteidigte und experimentell nahegelegte Abwägungsmodell eine Neuordnung des Machtverhältnisses zwischen Individuum und Kollektiv beziehungsweise Individuum und Staat. Kernpunkt dieser Ordnung ist eine Art „Kult der Gesellschaft“, der die Interessen des Ganzen über die Rechte des Einzelnen stellt. Die Implikationen, die ein solches Unterordnungsgebot enthält, sind nicht nur die Aufwertung der Quantität im Sinne des klassischen Weichensteller-Dilemmas, sondern, wie die Verrechnungskriterien der *Moral Machine* zeigen, zugleich die Diskriminierung in qualitativer Hinsicht. Im Opfer des Einzelnen gegenüber dem Kollektiv wird auch das Primat des Biologischen gegenüber der Ideologie geopfert.

Die politische Brisanz des Bruchs mit der Abwägungsresistenz der Menschenwürde wird deutlich in Analysen aktueller gesellschaftlicher Prozesse, die eine zunehmende Ermächtigung des Staates gegenüber seinen Bürgern konstatieren und gar vom „Kampf der Regierungen gegen die eigenen Bürger“ sprechen.<sup>30</sup> Die Rolle der neuen Technologien bei der Errichtung dieses Herrschaftsregimes liegt mit Blick auf ihr Potenzial einer umfassenden Überwachung (Dataveillance) im Kontext des „Internets der Dinge“ auf der Hand und gewinnt mit dem Sozialkredit-Modell Chinas eine neue Dimension. Dieses Modell ist der Vorschein eines kybernetischen Regierungssystems, das den Fokus von der moralischen Entwicklung des Individuums (und der Verinnerlichung moralischer Werte) auf das effektive Funktionieren des Systems (Netzwerk, Gesellschaft) verschiebt. Chinas Vorstoß darf allerdings nicht ablenken von vergleichbaren

---

28 H. Joas, *Die Sakralität der Person. Eine neue Genealogie der Menschenrechte*, Berlin 2011.

29 M. Neumann/T. Gutmann/T. Weitin/B. Fateh-Moghadam (Fn. 27), S. 138.

30 B. E. Harcourt (Fn. 24).

Prozessen im Westen im Kontext der digitalen Technologien. So ist das Individuum nur rhetorisch *Zweck* der Analyse seines Verhaltens im Internet, die ihm aus der besseren Kenntnis seiner Handlungen verbesserte Serviceleistungen verspricht. Aus einer philosophisch angelegten Perspektive, die den Menschen als ein sich selbst interpretierendes Wesen im Sinne der Persönlichkeitsentwicklung versteht, reduzieren auch solche Datenerhebungen das Individuum zum *Mittel*, nämlich der Erhöhung seiner eigenen Konsumbereitschaft.<sup>31</sup>

In dieser Hinsicht ist gerade auch die Rechtswissenschaft gefordert, praktizierte Verfahren zu bewerten und Regelungen zu schaffen, die weit über das Urheberrecht bezüglich Suchmaschinen, das Erbschaftsrecht bezüglich Facebook-Accounts oder das Eigentumsrecht bezüglich der im Prozess der Digitalisierung und Datafizierung produzierten und akkumulierten Daten hinausgehen. Die Diskussion muss so prinzipiell-politisch sein wie im Falle der Unantastbarkeit der Menschenwürde und zum einen fragen, wer die Algorithmen programmiert und beherrscht, zum anderen, wie sie die Machtverhältnisse und ökonomische Bedingungen in der Gesellschaft verändern und welche politischen Folgen dies hat. Das konkrete Resultat dieser Fragen könnte das Verbot bestimmter Forschungen einschließen, wenn deren mögliche Folgen als so problematisch und gefährlich für die Gesellschaft eingeschätzt werden, wie im Falle der Stammzellenforschung oder der autonomen Waffen. Dass auch im Bereich der KI-Forschung Risiken vermutet werden, vor denen es die Gesellschaft zu schützen gilt, zeigt nicht zuletzt die Mahnung von KI-Experten im Januar 2015: „Our AI systems must do what we want them to do.“<sup>32</sup>

Zugleich wird der Rechtsdiskurs auf die prinzipielle Herausforderung reagieren müssen, die in der Disposition der neuen Technologien liegt. „Die Konstruktionsweise und Funktion der Maschinen erleichtern oder ermöglichen bestimmte kognitive Prozesse, Handlungen oder Einstellungen und erschweren oder verhindern andere“, so formuliert die Technikphilosophin Catrin Misselhorn gut ein Jahrhundert nach Simmel das Phänomen der „Sachordnungen“, die sich der Mensch mit seinen Erfindungen schafft und denen er sich dann kaum entziehen kann.<sup>33</sup> Im Bereich der Maschinenethik, so heißt es weiter, besteht „die Gefahr, dass wir unsere Moralvorstellungen daran anpassen, was sich programmieren lässt.“<sup>34</sup>

---

31 S. Zuboff, *Das Zeitalter des Überwachungskapitalismus*, Berlin 2018.

32 *Future of Life Institute*, An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence, January 2015, <http://futureoflife.org/ai-open-letter>.

33 C. Misselhorn (Fn. 13), S. 126 f.

34 C. Misselhorn, S. 127.

Eine solche Anpassung wird nicht nur durch Entscheidungsmodelle wie die *Moral Machine* forciert, sie äußert sich noch im Bestreben, AI-Systeme als juristische Person („robot personhood“) zu etablieren, weil sich im komplexen Verfahren des „machine learning“ mögliche Fehlleistungen nicht auf konkrete Programmierschritte zurückführen lassen, um Verantwortlichkeiten zu rekonstruieren. Um die weitere Entwicklung und Anwendung von AI nicht am Black-Box-Problem scheitern zu lassen, wird daher vorgeschlagen, die Haftung für einen durch AI verursachten Schaden auf diese selbst zu übertragen und über eine für AI-Systeme gesetzlich vorgeschriebene Haftpflichtversicherung zu begleichen.<sup>35</sup> Ein weiteres Beispiel dafür, dass juristische Regelungen sich dem technisch Möglichen unterordnen, ist die mangelnde Auskunftspflicht in der EU-Datenschutz-Grundverordnung zum Zustandekommen algorithmischer Entscheidungen, die nicht nur aus Zwängen des Copyrights und Wettbewerbs resultiert, sondern auch aus Grenzen der Erklär- und Durchschaubarkeit algorithmischer Applikationen.<sup>36</sup>

Die hier unternommen Überlegungen führen vor Augen, dass mit künstlicher Intelligenz eine Technik vorliegt, die eine utilitaristische Ethik favorisiert. Algorithmen in autonomen Fahrzeugen sind schwer vereinbar mit dem im Deutschen Grundgesetz verankerten deontologischen Prinzip der unantastbaren Menschenwürde. Ihr Operationsmodus, der eine Sachordnung der Mathematik ist, subvertiert zwangsläufig die Unantastbarkeit der Menschenwürde. Folgt die Rechtsdogmatik diesem Ansinnen, wäre es

- 
- 35 J. Delker, Europe divided over robot ‘personhood’, Politico vom 11. April 2018, [www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood](http://www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood); zum Einspruch von mehr als 150 KI-Spezialisten vgl. den offenen Brief an die EU-Kommission AI and Robotics: Open Letter to the European Commission Artificial Intelligence and Robotics, 2018 <https://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf>.
- 36 J. Burrell, How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms, *Big Data & Society* 2016, S. 1–12. Der epistemische Einschnitt des Black-Box-Problems schlägt sich noch in den Lösungsvorschlägen nieder, wenn nicht Kausalitäten erstellt, sondern Korrelation erprobt werden: „Unlike existing approaches that try to provide insight into the internal logic of black box algorithms, counterfactual explanations do not attempt to clarify how decisions are made internally. Instead, they provide insight into which external facts could be different in order to arrive at a desired outcome“. S. Wachter/B. Mittelstadt/C. Russell, Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, *Harvard Journal of Law & Technology*, 31 (2018), S. 841–887, hier: S. 880. Für die praktische Umsetzung eines solchen Ansatzes vgl. Googles What-If Tool zur Inspektion von machine learning Modellen: *What-If Tool*, <https://pair-code.github.io/what-if-tool>.

ein Beispiel für die Rückkoppelungseffekte zwischen Objekt und Subjekt, Wirkung und Ursache oder eben Technik und Kultur und veranschaulichte auf bedrückende Weise die Geltung eines der prominentesten Sätze aus dem Denkkreis des Medienwissenschaftlers Marshall McLuhan auch im Bereich der Rechtswissenschaft: Wir formen unser Werkzeug und danach formt unser Werkzeug uns.

