

Transparenz als Mittel der Verantwortlichkeit bei KI-gestützten Systemen

Andreas Richter

Richter.AR@gmx.de

Abstract. Today, large IT companies provide users with many different complex algorithmic systems, while the users need to trust the former to handle their data responsibly. The trust in the providers is threatened, if the exact usage is not immediately apparent. I examine the question of accountability with a focus on transparency. Transparency as a means of accountability is not always feasible, desirable or appropriate. In my talk, I will address four limitations regarding the parties involved, the purpose, the object, and the possibility of transparency, based on, though expanding, the results of Ananny and Crawford (2018). Finally, two thought experiments are presented, which exemplify dilemmas of security and privacy on the one hand, and functionality and conflicts of interest on the other.

Keywords: Transparency, Accountability, AI Systems, IT Companies.

Zusammenfassung. Große IT-Unternehmen stellen Nutzern heutzutage komplexe algorithmische Systeme vielseitig zur Verfügung, während diese darauf vertrauen müssen, dass erstere verantwortungsvoll mit ihren Nutzungsdaten umgehen. Dabei droht dieses Vertrauen in die Anbieter zu erodieren, wenn die genaue Nutzung für Anwender nicht unmittelbar einsichtig ist. Ich untersuche hierzu die Frage nach Verantwortlichkeit und Rechenschaftspflicht (accountability) mit Fokus auf Transparenz. Transparenz als Mittel der Verantwortlichkeit ist nicht in jedem Fall umsetzbar, wünschenswert oder angemessen. In meinem Vortrag werde ich auf vier Begrenzungen eingehen, die sich hinsichtlich der Beteiligten, dem Zweck, dem Objekt und der Möglichkeit der Transparenz ergeben, wobei meine Analyse dieser Dimensionen an die Ergebnisse von Ananny und Crawford (2018) anknüpft. Abschließend werden zwei Gedankenexperimente vorgestellt, die modellhaft Dilemmata von Sicherheit und Privatsphäre, sowie von Funktionalität und Interessenskonflikten veranschaulichen.

Keywords: Transparenz, Verantwortlichkeit, Rechenschaftspflicht, KI-Systeme, IT-Unternehmen.

1 **Digitalkonzerne, Verantwortung und informationelle Selbstbestimmung**

Im Bereich der von Algorithmen gestützten Produkte und Dienstleistungen scheinen sich je Marktsegment nur wenige große oder gar nur ein Unternehmen an der globalen Spitze halten zu können. Unter Umständen kann eine marktdominierende, unternehmerische Macht bis zu einem gewissen Grad die Masse von Nutzern und Konsumenten beeinflussen, kontrollieren und damit ihre Autonomie einschränken. Welche Arten von Regeln kommen infrage, um den neuen Umständen und Herausforderungen für die Nutzerrechte zu begegnen? Ist eine einfache Forderung nach Transparenz zwischen den modernen Marktteilnehmern ausreichend, oder braucht es spezifischere Ansatzpunkte, um Verantwortung in Zeiten der Digitalisierung zuzuschreiben und um zu schützen, worauf es uns ankommt? Die Frage, der ich in diesem Kontext nachgehen werde, ist jene nach der Legitimation der Datenerhebung, der Rechenschaft wirtschaftlicher und politischer Akteure und dem Ideal der Transparenz als Mittel der Verantwortlichkeit.

Um uns dem Komplex der Verantwortlichkeit bei datenintensiven, KI-gestützten Systemen zu nähern, wenden wir uns zuerst Catrin Misselhorn zu, die in ihrem Werk „Maschinenethik“ der Frage nachgeht, inwiefern und wie man Moral in Maschinen implementieren könnte. Sie untersucht hierbei die Zuschreibung von Verantwortung in soziotechnischen Systemen, bei denen Menschen von algorithmischen Entscheidungsassistenten unterstützt werden. [vgl. 1, S. 126-135] Misselhorn stößt bei ihrer Untersuchung unter anderem auf das „*Problem der vielen Hände*“ [1, S. 132, verweist auf 2-4], nach dem die Verantwortungszuschreibung in der Softwareentwicklung durch den Umstand erschwert wird, dass nicht bestimmte, sondern eine Vielzahl von Akteuren am System als Endprodukt beteiligt waren, sodass einzelne Programmierer oft nicht zur Rechenschaft für Softwarefehler gezogen werden können, noch weniger die Endnutzer. Weiter können zum Teil nicht einmal die Entwickler vorhersagen, zu welchem Ergebnis KI-Systeme, die mit großen Datensätzen arbeiten, auf Basis der gegebenen Datenmenge kommen. Damit ist die Beurteilbarkeit der Entscheidungsempfehlung überhaupt gefährdet.

Misselhorn plädiert am Beispiel des autonomen und vernetzten Fahrens für den Schutz der informationellen Selbstbestimmung. Sie kritisiert den Schutz der Privatsphäre und der informationellen Selbstbestimmung von Seiten des deutschen Staates als unzureichend, da in dem Bericht der Ethikkommission zum autonomen und vernetzten Fahren Sparsamkeit in der Datenerhebung mit Gefahrenabwehr und wirtschaftlichen Interessen aufgewogen wird. [vgl. 1, S. 200-204, zitiert 5, S. 24] Außerdem hält sie die „Fülle und Zusammenführung der Informationen“, sowie „deren Verwendung“ [1, S. 201] grundsätzlich für problematisch für die informationelle Selbstbestimmung der Nutzer und plädiert für eine stärkere Betonung des Datenschutzes. Die Zukunft sieht sie aus diesen Gründen eher im assistierten als im autonomen Fahren. Somit entspricht bei Misselhorn folglich angemessener Datenschutz offenbar Datensparsamkeit, was schwer mit den Interessen von Digitalunternehmen vereinbar ist, bei denen die gesammelten und verarbeiteten Feedbackdaten der Nutzer selbst zur Wertschöpfung beitragen.

Eine bemerkenswerte Analyse der Situation und der Dynamiken auf dem Markt moderner, digitaler Produkte und Dienstleistungen legen Thomas Ramge und Viktor Mayer-Schönberger vor. Sie wenden unter anderem gegen die einseitige Betonung des Datenschutzes im Sinne der informationellen Selbstbestimmung ein, dass sie einem „Ungleichgewicht beim Zugang zu Daten“ [6, S. 204, vgl. hierzu 6, S. 203-205] Vorschub leistet, da Unternehmen solche Regelungen zum Teil umgehen können und Nutzer ihre Rechte oft durch einen unbedachten Klick abtreten. Stattdessen fordern sie „neue, bessere Transparenzregeln, die den zunehmenden Datenreichtum abbilden“ [6, S. 203] und eine Verschiebung des Fokus von der Datenerhebung auf die Datennutzung. Im Folgenden werde ich nachgehen, was mit dieser Forderung in etwa gemeint sein könnte.

2 Transparenz im Sinne von Verantwortlichkeit und Rechenschaftspflicht

Mike Ananny und Kate Crawford streben in ihrem Artikel zu den Grenzen des Transparenzideals im Sinne der Verantwortlichkeit und Rechenschaftspflicht an, die Grundlagen für die Frage danach bereitzustellen, welchen Zielen und Zwecken Forderungen nach und Maßnahmen der Transparenz eigentlich dienen sollen. [vgl. 7, S. 2-6&10-13] Ihnen zufolge müssen wir uns in Sachen der Transparenz bei soziotechnischen und algorithmischen Systemen immer die Frage stellen, „what is being looked at, what good comes from seeing it, and what are we not able to see?“ [7, S. 13] Laut den beiden Autoren wird im Allgemeinen angenommen, Transparenz sei ein Weg, auf dem man in die Wahrheit eines algorithmischen Systems Einsicht erhält und sei daher nötig, um Systeme verstehen und kontrollieren zu können, sowie der Rechenschaftspflicht nachzukommen und Verantwortung zuzuschreiben. Heutzutage fallen diesbezüglich simple Forderungen auf, Algorithmen doch uneingeschränkt offenzulegen. Die Autoren zeichnen die Spur des Ideals der Transparenz von seinen Ursprüngen in der Aufklärung bis zu heutigen computerwissenschaftlichen Debatten, in denen Transparenz als Einsicht in und damit Kontrolle über algorithmische Systeme gefordert wird.

Doch weil Sehen nicht gleich Wissen bedeutet, kann eine zu allgemeine Auffassung der Transparenz nach Ananny und Crawford ihren Anspruch als Mittel der Verantwortlichkeit oft nicht erfüllen. Daher ist bei algorithmischen und soziotechnischen Systemen der Blick in ihr Inneres unzulänglich und es braucht stattdessen einen Blick „across them – seeing them as sociotechnical systems that do not contain complexity but enact complexity by connecting to and intertwining with assemblages of humans and non-humans.“ [7, S. 2] Netzwerke aus Menschen und Maschinen müssen in ihren Beziehungen zueinander analysiert werden, sodass sich der Fokus der Transparenz spezifisch nach ihrer Absicht, ihrem Gegenstand und Ziel richtet und der Anspruch, Verantwortung zuzuschreiben, eher erfüllt wird, als mit einer einfachen Forderung zur Offenlegung aller Daten oder Algorithmen eines Systems.

Ananny und Crawford zählen zehn Begrenzungen des Transparenzideals auf, die zumindest die am meisten verbreiteten Probleme umfassen sollen. Sie schlagen zum

Abschluss ihres Artikels vor, diese wiederum zum Anfang der Diskussion zu nehmen und als Werkzeuge der Rechenschaftspflicht und der Zuschreibung von Verantwortung in algorithmischen und soziotechnischen Systemen zu nutzen. [vgl. hierzu im Folgenden 7, S. 6-10 & 12f.] Da die zehn Limitationen ohne weitere, dahinterliegende Struktur vorgestellt wurden, habe ich versucht, diese zehn Grenzen des Transparenzideals in vier wechselseitig aufeinander verwiesene Bereiche zu gruppieren, in denen Abwägungen nötig sind, um der Rechenschaftspflicht nachzukommen und um Verantwortung zuschreiben zu können. Diese sind nicht als absolut zu verstehen und hängen zum Teil voneinander ab. Der erste dreht sich um Macht und Interessensausgleich, der zweite um den Zweck der Maßnahmen in Bezug auf Signifikanz und Bewertbarkeit der transparent gemachten Informationen, der dritte Bereich zielt auf generelle Abwägungen zum Objekt der Transparenz und der vierte betrifft vor allem sich verändernde und unvorhersehbare Systeme.¹ Je nach Richtung und Kontext verweisen diese Aspekte der Transparenz zum Teil auf die Perspektive der Nutzerinteressen und auf die Legitimitätsfrage, sowie dementsprechend auf einen verantwortlichen Umgang mit der Erhebung und Nutzung von Daten allgemein.

2.1 Abwägungen hinsichtlich der involvierten Parteien

Erstens ergeben sich bei Forderungen nach Transparenz a) Probleme hinsichtlich der verschiedenen Akteure und dem Schutz ihrer jeweiligen Interessen. Da uneingeschränkte Transparenz sowohl für Einzelne in ihrer Privatsphäre als auch für Unternehmen in ihren Geheimnissen über die Grundlage ihrer Wertschöpfung schädlich sein kann, muss die Risiko- und Schadensverteilung unter den beteiligten Akteuren hinsichtlich ihrer Schutzbedürftigkeit untersucht werden. Wenn b) das Objekt der Transparenz in seiner Position gegenüber öffentlicher Zurschaustellung und Scham immun ist, bleibt Transparenz weiterhin folgenlos. Die Frage nach der Verantwortlichkeit in einem System muss also die Struktur von Machtgefällen thematisieren. Auf der einen Seite stehen somit Nutzer, die oft genug nicht völlig transparent „ausgeleuchtet“ werden wollen, auf der anderen Seite Unternehmen, die zum Teil ebenso ein legitimes Interesse daran haben, gewisse Informationen, zum Beispiel als Unternehmensgeheimnisse, zurückzubehalten. Gleichzeitig sind Unternehmen aber meist in der mächtigeren Position gegenüber dem einzelnen Nutzer ihrer Dienste und ihre Interessen müssen deswegen immer kritisch geprüft und auf ihre Legitimität hin hinterfragt werden.

2.2 Abwägungen hinsichtlich des Zwecks

Zweitens müssen die transparent gemachten Informationen entsprechend dem verfolgten Zweck einer gewissen Signifikanz, beziehungsweise einem Standard unterliegen, um bewertbar zu sein. Man könnte hier auch der Forderung nach einem transparenten Informationsmanagement sprechen. Unternehmen können a) mit der

¹ Wären die Grenzen ursprünglich von 1 bis 10 durchnummeriert, dann wäre die hier wiedergegebene Reihenfolge: (I) 2, 1, (II) 3, 5, 7, (III) 4, 6, (IV) 9, 10, 8.

Herausgabe von großen Massen an Informationen, die vor einer möglichen Bewertung aufwendig durchgearbeitet werden müssten, von wichtigen Details ablenken oder diese sogar verschleiern. Aus diesem Grund muss Ausschau gehalten werden nach Bemühungen darum, die Aufmerksamkeit im Informationsfluss und –überfluss undurchsichtig zu lenken. Das Ideal der Transparenz setzt b) die neoliberalen Fiktionen der rationalen Entscheidungsfindung und des Konsenses voraus, nach der alle Konsumenten alle ihnen zur Verfügung stehenden Systeme gleich einsehen und auch verstehen können, um sie zu vergleichen. Diese Fiktionen bleiben immer unerfüllte Fiktionen. Um dem Ideal nachzukommen, ist die Förderung der Motivation und der Befähigung zu dieser kollektiven Übersicht nötig.² Und wenn c) professionelle Expertise leitend sein soll, dann muss die Autorität, die die Deutungshoheit über verfügbare Informationen besitzt, angemessen hinterfragt werden. Mit ihrer Entwicklung und ihrem Einsatz muss verantwortlich umgegangen werden, um unlautere Forschung zu vermeiden, die für bestimmte Unternehmen günstige Zweifel und Meinungen säht. Man denke hier beispielsweise an die Stellungnahme der 100 Lungenärzte im Anschluss an den Abgasskandal oder die von Monsanto finanzierte Studie zur Schädlichkeit von Glyphosat, womöglich zu gegebenem Zeitpunkt in Zukunft auch an von Digitalriesen eigens gestifteten Lehrstühlen zur Ethik der künstlichen Intelligenz. Dieser Aspekt ist zum Teil mit dem ersten Bereich und der Frage nach Macht und Interessen verknüpft.

2.3 Abwägungen hinsichtlich des Objekts

Drittens ergeben sich objektspezifische Abwägungen hinsichtlich der Art und Intensität von Maßnahmen der Transparenz, sowie ihrem genauen Fokus, um falsche Einseitigkeiten und Misstrauen zu vermeiden. Auf der einen Seite sind a) je nach Absicht, die mit einer Forderung verfolgt wird, für das Objekt der Maßnahme die Art der Verantwortlichkeit und die Bedingungen für ihre Zuschreibung festzustellen. Statt sich zwischen absoluter Geheimhaltung und totaler Offenlegung zu entscheiden ist die Festlegung des angemessenen Grades und der genauen Objekte der Transparenz nötig. Dementsprechend muss erwogen werden, unter welchen Umständen eine Maßnahme der Transparenz je nach Objekt angemessen, passend oder überhaupt hilfreich sein kann, damit der Forderung bestmöglich entsprochen werden kann. Auf der anderen Seite ergeben sich b) zweiseitig Vertrauensfragen. Einerseits für externe Interessensgruppen, wie Anteilseigner eines Unternehmens oder auch Nutzer eines Dienstes, die einem System vertrauen, je nachdem wann welche Informationen zur Verfügung stehen und wie sie diese bewerten. Andererseits haben interne Akteure, Entwickler beispielsweise, teils Grund zur Befürchtung, jemand könne Daten, Wissen und Technik gezielt zum Schaden gebrauchen. Die Zweifel und Bedenken der Beteiligten sind damit von Bedeutung dafür, welche Teile des Systems bis zu welchem Grad für wen transparent sein sollten, da diese Objekte der Transparenz je besondere

² Dieser Aspekt ergänzt die Kernaussage von „Das Digital“, nach der der Markt im Datenkapitalismus „endlich sein volles Potential entfalten“ [6, S. 25] kann, um die Betonung der dazu nötigen Bedingungen und Befähigungen auf Seite der Konsumenten.

Eigenschaften der betrachteten Systeme oder auch Informationen über bestimmte Nutzer offenbaren können. Meines Erachtens nach rechtfertigt dies in Verbindung mit der ersten Gruppe von Abwägungen zum Beispiel die Forderung zur Kenntlichmachung von Chatbots als solche gegenüber dem Anwender, da dem Interesse an dieser Information auf Nutzerseite keine schutzbedürftigen Geschädigten gegenüberstehen, sondern nur rein wirtschaftliche Gesichtspunkte der Anbieter.

2.4 Abwägungen hinsichtlich der Möglichkeit

Schließlich nennen die Autoren viertens Probleme, die sich speziell bei selbstlernenden, sich verändernden Systeme stellen. Man muss hierbei dazusagen, dass diese Probleme je nach angewandter Methode des Selbst-Lernens variieren. Die Möglichkeit der transparenten Einsicht in solche Systeme ist a) technisch begrenzt. Sie sind teils so komplex, dass es selbst für Entwickler nicht ausreicht, ihren Code zu betrachten, um die Ursache für bestimmte Fehler zu finden und nachzuvollziehen, wie das System zu einem bestimmten Output kam oder welche Details die Berechnung wie stark beeinflusst haben. Bevor wir solche Systeme in verantwortlicher Position einsetzen, müssen wir sie ausreichend verstanden haben und erklären können und bis dahin weiterentwickeln und -erforschen, wenn wir sie überhaupt in unterschiedlichen Umgebungen und in entscheidenden Positionen einsetzen wollen. Dazu kommen b) zeitliche Beschränkungen der transparenten Beschreibbarkeit. Zum Nachvollzug eines bestimmten Outputs eines selbstlernenden Systems, zum Beispiel in Form einer Entscheidungsempfehlung, kann sogar dessen Quellcode und Set an Trainings- und Testdaten ungenügend sein, da durch den konstanten Lernprozess nie ein stabiler Zustand zu beobachten ist. Jede einzelne Berechnung beeinflusst potenziell jede weitere, daher ist das zukünftige Verhalten nicht sicher vorherzusehen und das vergangene Verhalten kann aus dem jetzigen Zustand womöglich nicht mehr erklärt werden. Mittels spezifischer Zeiträume und Rhythmen der Beobachtung müssen Wege gefunden werden, wie der Rechenschaftspflicht auch unter diesen erschwerenden Umständen nachgekommen werden kann. Zuletzt nenne ich c) das Problem, das zentral für den Artikel steht: „Seeing inside a system does not necessarily mean understanding its behavior or origins.“ [7, S. 8] Um komplexe Systeme und ihr Verhalten verstehen zu können, ist es nötig, dynamisch mit ihnen zu interagieren. Welche Eingaben und Veränderungen dabei möglich sind, ist schon eine bestimmende Vorauswahl für die Interaktion. Zentral ist es daher, die Art und Weise der Veränderung des Systems zu untersuchen, sowie die Akteure zu kritisieren, die in der Position sind, diesen Wandel zu steuern und mit Änderungen zu experimentieren.

3 Kurze Zusammenfassung unter Bezugnahme auf das autonome Fahren

Ich werde nun eine kurze Zusammenfassung zur Verantwortlichkeit und Rechenschaftspflicht formulieren und anschließend die Ergebnisse wie Misselhorn auf das paradigmatische Beispiel des autonomen Fahrens beziehen. Nach Ananny und

Crawford ist Transparenz als Mittel der Verantwortlichkeit nicht in jedem Fall umsetzbar, wünschenswert oder angemessen, stattdessen stellen sich mitunter Probleme hinsichtlich der Beteiligten, dem Zweck, dem Objekt und der Möglichkeit der Transparenz. Die zehn Begrenzungen des Transparenzideals habe ich zusammengefasst in vier wechselseitig verschränkte Bereiche, in denen Abwägungen nötig sind, um der Rechenschaftspflicht nachzukommen und um Verantwortung zuschreiben zu können. Wie bemerkt lassen sich diese Problembereiche teils als Verweise auf einen verantwortlichen Umgang mit Daten verstehen, der Nutzerinteressen und -rechte respektiert.

Erstens sind Abwägungen erforderlich zu den involvierten Parteien, wobei im Bereich der Automobilität alle Machtverhältnisse in Unternehmen und Politik und die Schutzbedürftigkeit und Interessen aller beteiligten Gruppen analysiert und adressiert werden müssen. Zweitens müssen die Maßnahmen der Transparenz ihrem Zweck nach ausgerichtet sein. Im Sinne eines transparenten Informationsmanagements sollten die veröffentlichten Informationen zu Herstellung und Programmierung, Funktion, sowie Bedienbarkeit der autonomen Fahrzeuge und Fahrdienste hinsichtlich der Ziele der Forderung nach Transparenz signifikant, für den Rezipienten zu überblicken und unabhängig bewertbar sein. Drittens sind der genaue Grad, das Ziel und die Intensität der Transparenz auf ihr Objekt, hier vertrauenswürdige und sichere, selbstfahrende Automobile und Fahrdienste, abzustimmen. Beispielsweise kann die detaillierte Veröffentlichung von Funktionsweisen, Quellcodes und Schwachstellen im schlimmsten Fall Manipulation und Schädigung erlauben. Und schließlich müssen viertens die Besonderheiten sich verändernder, selbstlernender KI-Systeme speziell berücksichtigt werden, wobei sich vor allem zeitliche und technische Probleme stellen, die periodische Beschreibungen, sowie weitere Forschung und Entwicklung nötig machen, als auch neue Fragen der Verantwortlichkeit und Rechenschaftspflicht für unerwünschte Fehlfunktionen und Unfälle aufwerfen.

Beim autonomen Fahren als beispielhaftem Fall der Automatisierung menschlicher Tätigkeit stellen sich die Fragen, wie viel Vernetzung, Kontrolle, und Transparenz der Daten nötig, schädlich oder aus moralischer Perspektive wünschenswert ist. Insgesamt werden wir aufgefordert sein, darüber nachzudenken, welche Daten wir in öffentlicher Hand brauchen und welche wir privat halten wollen, welche Daten Unternehmen unter welchen Umständen wie nutzen und wem zur Verfügung stellen dürfen oder auch müssen, wem personenbezogene Daten gehören, wie man sensible Daten sicher vor unautorisiertem Zugriff schützt und wie man sie standardmäßig und zuverlässig anonymisiert, sodass ein Rückvollzug auf Einzelpersonen nicht möglich ist. Wir müssen uns paradigmatisch im Straßenverkehr fragen, bei welcher Datennutzung die Sicherheit aller, mitunter vermittelt durch die Sicherheit der Produkte, im Vordergrund stehen muss und mehr Datenerfassung zu höherer Sicherheit führt und bei welcher Nutzung die Autonomie des Einzelnen, seine informationelle Selbstbestimmung und seine Privatsphäre eine übergeordnete Rolle spielt und daher seine Zustimmung notwendig ist. Damit verbunden ist ein Bewusstseinswandel in Bezug auf die Datenerhebung und -nutzung, der möglicherweise notwendig macht, dass wir Privatsphäre und Autonomie des Einzelnen, sowie Nutzerrechte und informationelle Selbstbestimmung im Sinne der Hoheit über personenbezogene Daten in einer

digitalisierten, demokratischen Gesellschaft zum Teil neu definieren müssen. Diesen Fragestellungen und ethischen Erwägungen werde ich im Folgenden noch in Form von Gedankenexperimenten nachgehen, bevor ich diesen Vortrag abschlieÙe.

4 Gedankenexperimente zur Datenerhebung und -nutzung

Die Probleme, auf die wir beim autonomen und vernetzten Fahren stoÙen, betreffen somit weit mehr Aspekte als in den berÙhmten Gedankenexperimenten der sogenannten Trolley Cases [sh. hierzu z.B. 1, S. 190-204, sowie 8-9] Eingang finden: die Anzahl geretteter und getöteter oder verletzter Menschen in Unfallsituationen und deren Lebenserwartung. Diese Fälle sind nicht kongruent mit Standardsituationen des Verkehrs. Es geht im Straßenverkehr nicht ständig und vor allem um Situationen, in denen man zwischen verschiedenen Gruppen von Menschen wählen muss, die man umfährt. Die betrachteten Fragen der Datenerhebung und -nutzung betreffen Abwägungen der Persönlichkeitsrechte und der Sicherheits- und Wirtschaftspolitik gleichermaßen, was die Besonderheit in diesem Feld ausmacht. Abwägungen zu Schutz und Nutzen sind elementar, aber auf der allgemeinen Ebene der zugrundeliegenden Intuitionen und Prinzipien, der praktischen Verfahrensweisen und rechtlichen Regelungen und nicht auf der der Konsequenzen einzelner autonomer Entscheidungen, wie es uns die Trolley Cases weiß machen wollen. Um für Klarheit zu sorgen, sollen an dieser Stelle zwei alternativ gelagerte Gedankenexperimente helfen, die sicherheitspolitische und wirtschaftspolitische Intuitionen reflektieren.

Erstens ein sicherheitspolitisches Gedankenexperiment, das ein Argument gegen vollkommen uneingeschränkte Privatsphäre verbildlichen soll. Stellen wir uns hierfür im beispielhaften Zusammenhang des Straßenverkehrs ein Szenario vor, in dem eine Person³ alle möglichen Daten zu einer Verkehrssituation erhält und alle Vorgänge verfolgt, aber niemanden aus der betrachteten Situation persönlich kennt und mit niemandem sozial interagiert. Dieser Beobachter schlägt einzig und allein dann Alarm oder greift ein, wenn Schutz und Wohl von Beteiligten oder Unbeteiligten gefährdet sind und kollektiv bestimmte und transparent verfolgte Normen verletzt werden. In diesem Modell stellt sich die Alternative ein, sich für oder gegen Datenerhebung und damit gleichzeitig für oder gegen ein erhöhtes Maß an Sicherheit zu entscheiden, wobei die Bewertung der Legitimität des Agierens dieser imaginierten Person entweder der Sicherheit oder der Privatsphäre und individuellen Autonomie mehr Gewicht gibt. Um dieses Gedankenspiel auf ein konkretes und drastisches Beispiel im Kontext zu bringen: Sollten wir es modernen Autos als Werkzeug der Fortbewegung durch einen Algorithmus verbieten, den Willen des Fahrers zu befolgen, wenn dieser im Begriff ist, es vorsätzlich als Waffe einzusetzen? Hier stellt sich die Vermutung ein, dass eine bestimmte Handhabung der Daten unter Umständen das Potential haben könnte, die Sicherheit im Sinne der körperlichen Unversehrtheit besser schützen zu können als bisher, ohne notwendig die Privatsphäre unverhältnismäßig einzuschränken.

³ Eine Person zur Vereinfachung, ohne Fragen zum Bewusstsein bei algorithmischen Systemen zu adressieren.

Zweitens will ich in einem allgemeineren Gedankenexperiment die individuelle Situation Einzelner veranschaulichen, um Unternehmensinteressen und die Interessen von privaten Konsumenten gegenüberzustellen, ohne Bezug zum Straßenverkehr. Hierzu schlage ich ein Szenario vor, bei dem Privatpersonen ständig einen Ratgeber in ihrer unmittelbaren Nähe haben, der kostenlos für jede Frage bereitsteht, allerdings auch immer und überall mithört. Gleichzeitig vertritt dieser Ratgeber die Interessen eines Kaufmanns, der für Produkte und Dienstleistungen wirbt. Er versucht zu jedem Zeitpunkt aus den Informationen, die er erhält, Rückschlüsse auf die Präferenzen seiner Nutzer zu ziehen und sie zur Erhöhung der Konsumbereitschaft in Bezug auf bestimmte Produkte zu verwenden. Damit besteht beim Dienst des Ratgebers stets ein Interessenskonflikt, wenn der Maßstab der Unabhängigkeit angelegt wird. In diesem zweiten Gedankenexperiment droht die Manipulation der Nutzer, wenn der Ratgeber nur entweder zu jeder Zeit zur Verfügung steht und anwesend ist oder gar nicht. Ein uneingeschränkter Zugang zu allen Informationen über eine Person gibt die Macht zur Beeinflussung, zum Beispiel durch subtile, persönlich angepasste Werbung, wodurch sich die Notwendigkeit darstellt, den Einzelnen die Macht über ihre Daten zu geben, sprich im Szenario gesprochen, die Nutzer darüber entscheiden zu lassen, was der Ratgeber von ihnen weiß und wann er zugegen ist oder zuhört. Außerdem sollten Interessenskonflikte vermieden und Werbung als solche erkenntlich sein. Ersetzt man übrigens die Kaufmannsinteressen mit denen eines ideologisch totalitär ausgerichteten Überwachungsstaats, der fanatisch das Ziel von Einheit oder Sicherheit verfolgt, so ergibt sich ein starkes demokratisches Regulativ zum ersten Gedankenexperiment bezüglich dem Einsatz staatlicher Kontrolle.

5 Datenmonopole und Demokratie

Ich habe in diesem Vortrag versucht, Verantwortlichkeit als Transparenz im Bereich algorithmischer und soziotechnischer KI-Systeme zu untersuchen. Dabei bin ich letztlich zu möglichen Auswirkungen auf die Autonomie und Privatsphäre des Einzelnen gelangt, die die Konzentration von Datenströmen und Entscheidungsmacht auf nicht vertrauenswürdige, im Sinne von intransparent agierenden Unternehmen haben können. Datenmonopole beeinträchtigen durch ihre Stellung und durch die damit mögliche Manipulation die Entscheidungsfreiheit ihrer Nutzer, deswegen sollten letztere durch effektiv durchsetzbare Rechte gegenüber mächtigen Digitalkonzernen geschützt sein und Zugang zu einer Auswahl von sich unterscheidenden Diensten und Informationsquellen haben. Macht, Überwachung und Kontrolle sind einzig zum Schutz von gemeinsam bestimmten Gütern und vor anerkannten Gefahren zu etablieren.

Monopol-, Quasimonopol- und Oligopolstellungen haben negative Auswirkungen, die sich in dem betrachteten Komplex der Transparenz finden lassen. Datenmonopole sind Machtkonzentrationen und diese müssen klare Strukturen der Verantwortlichkeit aufweisen. Transparenzregeln müssen einen Interessensausgleich zwischen Nutzer- und Unternehmensinteressen schaffen, bestimmte Informationen müssen unterschiedlichen Interessensgruppen zur Verfügung stehen und der Umgang mit

personenbezogenen Daten muss nachvollziehbar, vertrauensvoll und durch die Nutzer beeinflussbar sein, wobei die Besonderheiten von KI-gestützten Systemen besondere Darstellungen und Beschreibungen, sowie Vorsicht und intensive Forschungsbemühungen nötig machen.

Die Regeln, die den marktwirtschaftlichen Wettbewerb zwischen den Anbietern fördern sollen, sind dazu da, um aus Nutzer- und Konsumentensicht effizient für einen gleichen Zugang zu Gemeingütern und für deren Schutz zu sorgen. Das Ziel, welche Gemeingüter in welchem Verhältnis zueinander geschützt werden sollen, und die entsprechenden Regeln geben wir uns dabei in einer Demokratie dem Anspruch nach selbst.

Literaturverzeichnis

1. Misselhorn, C.: Grundfragen der Maschinenethik. Reclam (Reclams Universal-Bibliothek, Nr. 19583), Ditzingen (2018)
2. Jonas, H.: Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation. Insel-Verlag, Frankfurt am Main (1979)
3. Friedmann, B.: Moral Responsibility and Computer Technology. Paper presented at the annual meeting of the American Educational Research Association, Boston (1990)
4. Nissenbaum, H.: Computing and Accountability. *Communications of the ACM* 37 (1), 72–80 (1994)
5. Di Fabio, U., Broy, M., Jungo Brüngger, R. et al.: Ethik-Kommission. Automatisiertes und vernetztes Fahren. Bericht (2017), online verfügbar unter https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile (zuletzt geprüft: 12.01.2020)
6. Mayer-Schönberger, V., Ramge, T.: Das Digital. Markt, Wertschöpfung und Gerechtigkeit im Datenkapitalismus. 2. Aufl., Econ, Berlin (2017)
7. Ananny, M., Crawford, K.: Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20 (issue: 3), 973–989 (1981), online verfügbar unter <https://doi.org/10.1177/1461444816676645> (zuletzt geprüft: 12.01.2020)
8. Foot, P.: The Problem of Abortion and the Doctrine of Double Effect. In: Foot, P.: *Virtues and vices and other essays in moral philosophy*. Clarendon Press, Oxford (2002)
9. Thompson, J. J.: The Trolley Problem. *The Yale Law Journal* 94 (6), 1395–1415 (1985), online verfügbar unter <https://doi.org/10.2307/796133> (zuletzt geprüft am 12.01.2020)