

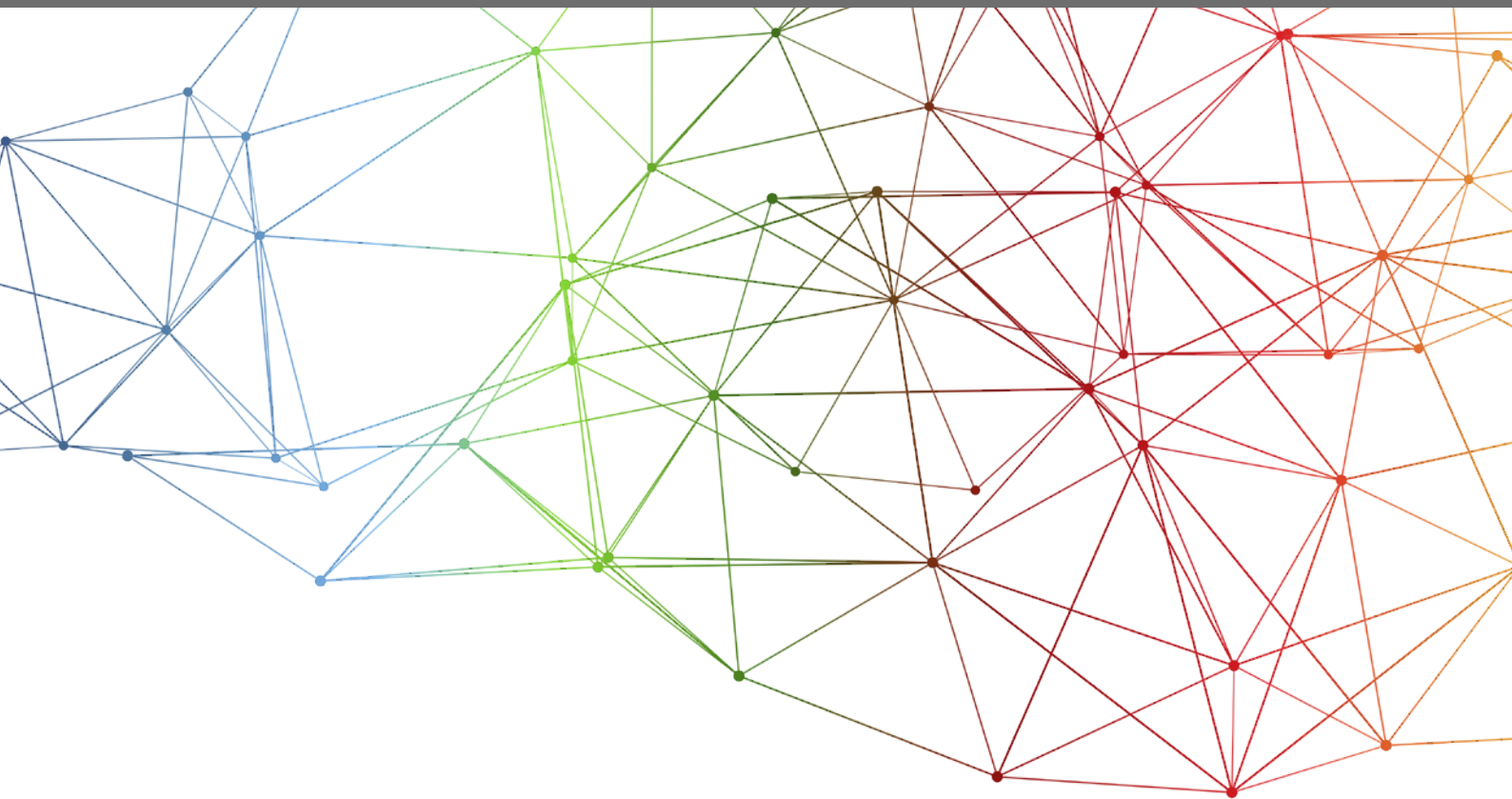


ALGORITHM
WATCH

AlgorithmWatch Arbeitspapier Nr. 4

Ethik und algorithmische Prozesse zur Entscheidungsfindung oder -vorbereitung

von Lorena Jaume-Palasi und Matthias Spielkamp



Zusammenfassung

Von Algorithmen gesteuerte, automatisierte Entscheidungsprozesse (automated decision-making, ADM) nehmen schon heute breiten Raum in unserer Gesellschaft ein. Diese Entwicklung steht gerade erst an ihrem Anfang. In wenigen Jahren werden alle Bürgerinnen und Bürger täglich auf die eine oder andere Art von Entscheidungen dieser Systeme betroffen sein. Automatisierung wird alle politischen und rechtlichen Felder erfassen.

Die gegenwärtige ethische Debatte um die Folgen von Automatisierung fokussiert auf individualrechtliche Aspekte. Jedoch weisen algorithmische Prozesse, die die Automatisierung vor allem ausmachen, primär eine kollektive Dimension auf. Diese kann nur partiell individualrechtlich adressiert werden. Die bestehenden ethischen und rechtlichen Kriterien sind aus diesem Grund für eine allgemeine Betrachtung nicht geeignet oder zumindest unzulänglich. Sie führen daher zu begrifflichen Vermischungen, wie etwa bei den Themen Privatheit und Diskriminierung, indem Informationen, die potentiell für illegitime Diskriminierung missbraucht werden können, als private Informationen deklariert werden.

Unser Beitrag versucht hier zum einen, Ordnung in die Debatte zu bringen, damit diese Vermischungen in Zukunft vermieden werden können. Zum anderen befassen wir uns mit ethischen Kriterien der Technik, die als allgemeine abstrakte Prinzipien auf alle gesellschaftlichen Kontexte anzuwenden sind.

Allerdings ist die Frage der Ethik immer eine Frage des Handelns und der Verantwortung für dieses Handeln – und folglich unabdingbar von einer strukturellen und situativen Kontextualisierung abhängig: So können die Regeln, die für den Staat gelten, kaum für den Bürger gelten. Diese Differenzierung, die klassisch in der Ethik und im Konstitutionalismus verwendet werden, fehlt bislang in der Debatte über Automatisierung.

Das vorliegende Arbeitspapier will mit einer Taxonomie einen Beitrag für eine differenzierte ethische und rechtliche Debatte leisten. Diese Taxonomie ist sowohl auf die Frage des Handelns, als auch auf die Frage der Dimensionen potentieller Schäden in der Automatisierung fokussiert.

Zunächst wird der Bedarf für eine technikneutrale Ethik begründet. Darauf folgend wird die Frage des Handelns und Entscheidens in der Automatisierung betrachtet und schließlich eine Taxonomie vorgeschlagen, die als Struktur die verschiedenen Risiken und Konflikte adäquater einordnet und damit eine differenzierte Entwicklung ethischer Kriterien ermöglicht. Als Kategorien schlagen wir *Individualgüter* und *gesellschaftliche Güter* vor, wobei gesellschaftliche Güter wiederum unterteilt sind in den *gesellschaftlichen Rahmen* und *Kollektivgüter*.

Diese Taxonomie strukturiert das Öffentliche als Dimension in seiner Gänze als komplexe Struktur, die sich nicht auf Meinungen und Informationen reduzieren lässt, sondern einerseits Kollektivgüter erfasst und andererseits individuelle und kollektive Interaktionen betrachtet. Somit wird eine bessere ethische Kontextualisierung für die Ausarbeitung differenzierter ethischer Kriterien angeboten, die eine technikneutrale, an Werten orientierte Betrachtung wieder in Mittelpunkt stellt.

Die Entstehung dieses Arbeitspapiers wurde durch ein Fellowship des Bucerius Labs der ZEIT-Stiftung Ebelin und Gerd Bucerius ermöglicht.



Inhalt

/	Einleitung: Ethik in der Digitalisierung	4
/	Über die Notwendigkeit einer technologieneutralen Ethik bei Algorithmen	5
	Ethik als programmierfähige Handlungsanleitung	5
	Technisch orientierte ethische Kriterien	5
	Ethik der erkenntnisbezogenen Einschränkung	6
/	Über die Kodierbarkeit von Entscheidungen und Ethik	9
	Menschliches Entscheiden und Verantwortung	9
	Rationalität und Handlungsfreiheit in der Automatisierung	10
/	Ethik im Raum: Eine Taxonomie des Gesellschaftlichen im automatisierten Zeitalter	12
	Gesellschaftliche Güter	13
	a) Der gesellschaftliche Rahmen	13
	b) Kollektivgüter	14
	Individualgüter	14
/	Ethik und struktureller Kontext. Eine Zusammenfassung	16
/	Literaturverzeichnis	18
/	Autorinnen	19
/	Impressum	20

Einleitung: Ethik in der Digitalisierung

Von Algorithmen gesteuerte, automatisierte Entscheidungsprozesse (automated decision-making, ADM) nehmen schon heute breiten Raum in unserer Gesellschaft ein. Sie reichen von Fahrerassistenz-Systemen, die Autos bei Gefahr abbremsen, bis hin zu Software, die darüber entscheidet, ob jemandem ein Kredit gewährt wird. Auch staatliches Handeln wird immer öfter von ADM-Systemen unterstützt, sei es in der „vorausschauenden Polizeiarbeit“ (predictive policing) oder bei der Entscheidung darüber, ob jemand aus dem Gefängnis entlassen wird. Diese Entwicklung steht gerade erst an ihrem Anfang. In wenigen Jahren werden alle Bürgerinnen und Bürger täglich auf die eine oder andere Art von Entscheidungen dieser Systeme betroffen sein. Automatisierung wird alle politischen und rechtlichen Felder erfassen.

Die gegenwärtige ethische Debatte um die Folgen von Automatisierung fokussiert auf individualrechtliche Aspekte. Jedoch weisen algorithmische Prozesse, die die Automatisierung vor allem ausmachen, primär eine kollektive Dimension auf. Diese kann nur partiell individualrechtlich adressiert werden. Die bestehenden ethischen und rechtlichen Kriterien sind aus diesem Grund für eine allgemeine Betrachtung nicht geeignet oder zumindest unzulänglich. Sie führen daher zu begrifflichen Vermischungen, wie etwa bei den Themen Privatheit und Diskriminierung, indem Informationen, die potentiell für illegitime Diskriminierung missbraucht werden können, als private Informationen deklariert werden. Unser Beitrag versucht hier zum einen, Ordnung in die Debatte zu bringen, damit diese Vermischungen in Zukunft vermieden werden können. Zum anderen befassen wir uns mit ethischen Kriterien der Technik, die als allgemeine abstrakte Prinzipien auf alle gesellschaftlichen Kontexte anzuwenden sind.

Allerdings ist die Frage der Ethik immer eine Frage des Handelns und der Verantwortung für dieses Handeln – und folglich unabdingbar von einer strukturellen und situativen Kontextualisierung abhängig: So können die Regeln, die für den Staat gelten, kaum für den Bürger gelten. Diese Differenzierung, die klassisch in der Ethik und im Konstitutionalismus verwendet werden, fehlt bislang in der Debatte über Automatisierung. Das vorliegende Paper will mit einer Taxonomie einen Beitrag für eine differenzierte ethische und rechtliche Debatte leisten. Diese Taxonomie ist sowohl auf die Frage des Handelns, als auch auf die Frage der Dimensionen potentieller Schäden in der Automatisierung fokussiert. Zunächst wird der Bedarf für eine technikneutrale Ethik begründet. Darauf folgend wird die Frage des Handelns und Entscheidens in der Automatisierung betrachtet und schließlich eine Taxonomie vorgeschlagen, die als Struktur die verschiedenen Risiken und Konflikte adäquater einordnet und damit eine differenzierte Entwicklung ethischer Kriterien ermöglicht.

Über die Notwendigkeit einer technologie-neutralen Ethik bei Algorithmen

Seit der zweiten Hälfte der 2000er Jahre ist die Frage der Ethik in Bezug auf Algorithmen und weitere Automatisierungsprozesse vermehrt Gegenstand der Forschung. Einige der Forschungsarbeiten (Bozdog, 2013; Naik & Bhide, 2014; Friedman & Nissenbaum, 1996; Tene & Polonetsky, 2013) blicken vor allem beschreibend auf die immanente Subjektivität, die der Programmierung von Algorithmen innewohnt. *Machine bias* – sprich Voreingenommenheit im Code –, das machen diese Forschungsarbeiten deutlich, ist ein unvermeidbares Ergebnis des kulturellen Hintergrunds und der Sozialisierung der Entwickler und Datenwissenschaftler, die entsprechende algorithmische Prozesse entwerfen und umsetzen.

Bereits 1996 entwickelten Batya Friedman und Helen Nissenbaum eine Typologie zu Bias in Computersystemen, die verschiedene Möglichkeiten der Verankerung von menschlichem Bias in maschinellen Prozessen beschrieb: „Bias can enter a [computer] system either through the explicit and conscious efforts of individuals or institutions, or implicitly and unconsciously, even in spite of the best of intentions“. Die Art und Weise, in der Diskriminierung durch die Verankerung in einem maschinellen System standardisiert werden, liegt der Typologie von Friedman und Nissenbaum zufolge nicht allein bei dem Entwickler oder Auftraggeber. Bias kann auch aus dem Kontakt mit dem Nutzer entstehen oder durch Konflikte in der Formalisierung sozialer Phänomene, die sich schwer in Code formulieren lassen. Aus ihrer Beobachtung, dass Bias und Diskriminierung sich in den maschinellen Prozessen selbst manifestieren, folgern Friedman und Nissenbaum, dass auch die ethische Analyse von ADM-Systemen vor allem auf technologischer Ebene ansetzen sollte.

Diese Grundintention wird von der überwiegenden Mehrzahl der Forschungsarbeiten geteilt, die sich mit dem ethischen Umgang mit Bias und Diskriminierungen im Zeitalter der Automatisierung befassen. Wir stellen im Folgenden drei Varianten der Konkretisierung des technologiebezogenen Ansatzes vor. Dabei zeigen wir auch auf, wo dieser Ansatz aus unserer Sicht deutliche Schwächen

aufweist oder implizit von sehr voraussetzungsreichen Annahmen ausgeht.

Ethik als programmierfähige Handlungsanleitung

Mit der Komplexität des Algorithmus steigt seine Undurchsichtigkeit: Bei manchen Arten selbstlernender Algorithmen können die Prozesse, die der Algorithmus entwickelt, um bestimmte Resultate zu erzeugen, selbst von ihren Entwicklern nicht erklärt werden. Aus diesem Grund werden Algorithmen einer ersten statt zweiten Ordnung (einer Meta-Ebene) in Betracht gezogen, die als ethische Instanz über dritte Algorithmen „wachen“ sollen (Etzioni, Turilli & Wiltshire, 2016; Anderson & Anderson, 2007). Dieser Vorschlag ist voraussetzungsreich: Er setzt voraus, dass ethisches Handeln in einer logischen Sprache programmiert werden und automatisiert werden kann – oder anders ausgedrückt: dass Algorithmen in der Lage sind, Gedanken, Erwägungen und letztlich Handlungen zu vollziehen. Weiter wird angenommen, dass eine Art des ethischen Programmierens ohne *machine bias* möglich sei. Bello und Bringsjord (2012), die in ähnliche Richtung argumentieren, räumen zwar ein, dass moralisches Denken bei Algorithmen nicht anhand klassischer ethischer Prinzipien strukturiert werden dürfe, da dieses moralische Denken bei Algorithmen nicht die Art reflektiere, in der Menschen ihre treffen. Sie bleiben jedoch eine Erklärung schuldig, inwiefern Algorithmen ‚denken‘ können und warum dieses algorithmische ‚Denken‘ als moralisch bezeichnet werden darf.

Technisch orientierte ethische Kriterien

Ein und derselbe Algorithmus kann sehr unterschiedlichen Zwecken dienen: ein Algorithmus zur Filmauswahl kann ebenfalls in der Krebsforschung nützlich sein. Die Funktionalität von Algorithmen muss daher im Kontext betrachtet werden. Sowohl Datenauswahl und Datenbasis, als auch der Anwendungsbereich spielen eine Rolle hinsichtlich der Risiken und Chancen eines algorithmischen Prozesses, insbesondere bei komplexeren Algorithmen.

Auf diesem Grund konzentrieren sich andere Ansätze auf die Erarbeitung von normativen Kriterien bei der Konzeptualisierung und Datenverarbeitung in algorithmischen Prozessen. Viele dieser Kriterien entstammen der Datenschutzkultur (Romei & Ruggieri, 2014; Kamishima, 2012), wie etwa die Zweckbindung. Sie besagt, dass die Datenverarbeitung an einen klar formulierten Zweck gebunden sein solle und, bei Verwendung personenbezogener Daten, auf einer gesetzlichen Grundlage oder der Einwilligung der Betroffenen fußen müsse. Enge Zweckbindungen sowie weitere Grundsätze wie das Prinzip der Datensparsamkeit führen jedoch dazu, dass bestimmte Korrelationen oder Suchkriterien nach Mustern praktisch ausgeschlossen werden, weil sie gegen Datenschutzregeln verstoßen.

Als Korrektiv wird häufig Transparenz gefordert: mal als Offenlegung des Codes (Tutt, 2016), mal als die Pflicht über die Art zu informieren, in der Daten algorithmisch verarbeitet werden (Datta et al., 2016; Tene and Polonetsky, 2013a). Diese Forderung nach Transparenz wird als *conditio sine qua non* gesehen, um eine sogenannte informationelle Privatheit selbstbestimmt bewahren zu können. Schermer (2011) geht sogar darüber hinaus. Er lehnt das Konzept der informationellen Privatheit ab, verlangt jedoch ein Recht auf Privatheit von Kollektiven, da Profiling die Identifizierbarkeit von Individuen irrelevant mache. Vielmehr versuche Profiling, Individuen in aussagekräftigen Gruppen zusammenzustellen, sodass die Identität der Individuen keine Relevanz mehr habe (Floridi, 2012; Hildebrandt, 2011; Leese, 2014).

Allerdings sind personenbezogene Daten nicht per se private Daten. Das Konzept der informationellen Privatheit reduziert die Vorstellung von Privatheit auf die reine Kontrolle von personenbezogenen Informationen und übersieht, dass eine Kontrolle der personenbezogenen Daten möglich ist, ohne damit Privatheit zu genießen. Wesentliches Merkmal des Privaten ist hingegen die Autonomie von politischen und ökonomischen Diktaten.

In der liberalen Tradition, die sowohl die öffentliche Diskussion wie auch die Rechtsprechung in Deutschland und der Europäischen Union maßgeblich prägt, wird Privatheit allgemein als Ermöglichungsbedingung und auch als Ausdruck von Autonomie im Sinne selbständigen Denkens und Handelns verstanden. (Wehofsits, 2016)

Kontrolle über die eigenen personenbezogenen, privaten Daten zu haben, ist lediglich eines von verschiedenen möglichen Mitteln, um Autonomie zu erlangen, aber kein Zweck an sich. (Davon abgesehen: Individuen hatten vor

den Zeiten der Digitalisierung keine absolute Kontrolle über ihre personenbezogenen Daten und haben diese Kontrolle auch heute nicht. Der Mensch ist ein soziales Wesen und teilt – mit oder ohne Absicht – personenbezogene Daten mit seiner Umgebung.)

Genau daran, dass die Kontrolle von personenbezogenen Daten absolut gesetzt wird, krankt die Idee von normativen Kriterien, die beim Algorithmen-Design oder der Datenverarbeitung ansetzen: Diesen Kriterien wird ein ethischer Charakter beigemessen, obwohl sie lediglich technische Einschränkungen entweder im Design oder im Verarbeitungsprozess sind – und zwar völlig vom Kontext gelöst. Diese Kriterien beziehen sich jedoch nicht auf menschliches Handeln im Sinne des Ethischen, sondern auf die Konzeptualisierung und Durchführung automatisierter Datenverarbeitung. Sie sind keine Handlungsanleitung an Entwickler oder Datenwissenschaftler, sondern an das Programm an sich. Programme handeln aber nicht, sie führen Aufgaben aus. Diese Programme sind entworfen und geprägt von der Welt seiner Designer und vom Individuum, das diese einsetzt. Ethische Anforderungen jedoch richten sich an Akteure, die im eigentlichen Sinne in der Lage sind, zu handeln.

Ethik der erkenntnisbezogenen Einschränkung

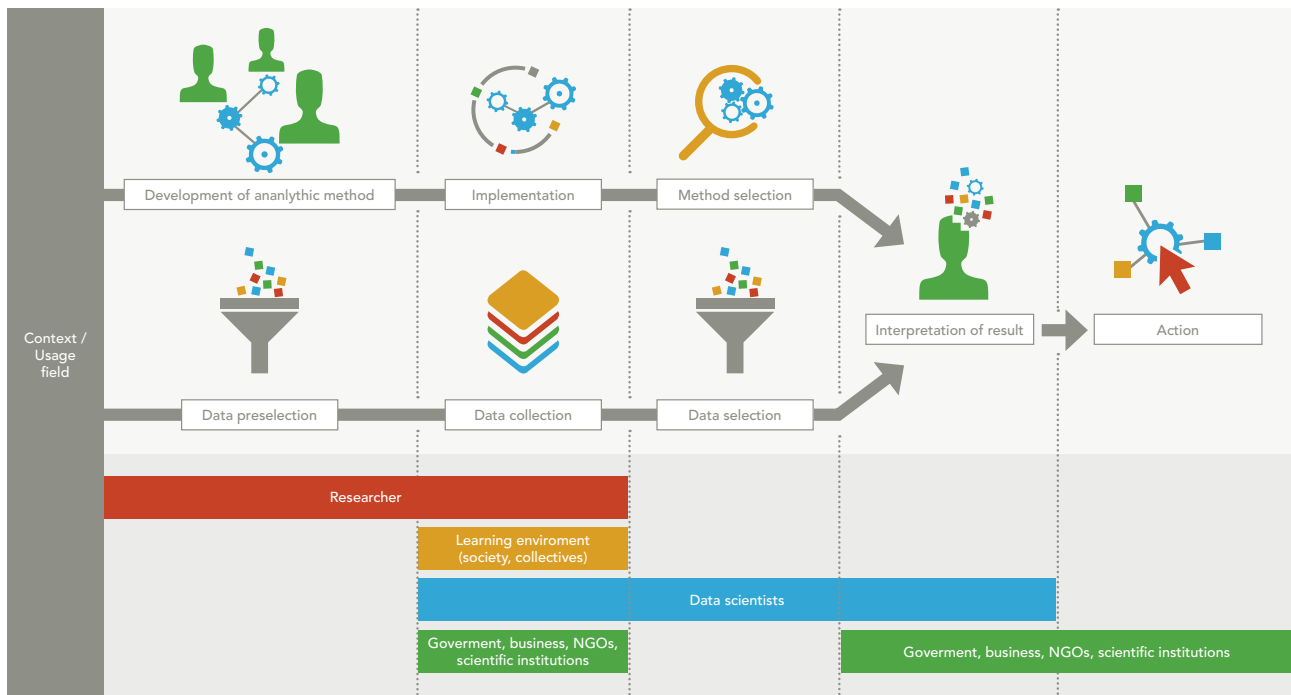
Ein dritter Strang verfolgt die Strategie, in den Erkenntnisprozess, der ADM-Systemen zugrunde liegt, einschränkend einzugreifen (Pasquale, 2015). Dieser Eingriff kann zum Beispiel durch den Ausschluss bestimmter Datenkategorien von der Datenverarbeitung erfolgen. So könnte man als Regel festlegen, dass etwa beim Scoring Gesundheitsdaten nicht mit Finanzdaten kombiniert werden dürfen, oder bei Personalauswahl oder der Vergabe von Mietwohnungen die ethnische Herkunft keine Rolle spielen darf. Mittelstadt et al. sehen ebenfalls Diskriminierungsrisiken nur auf Erkenntnisebene. Hier werden ethische Konflikte ausschließlich im analytischen Verfahren komplexer Algorithmen verortet. Entweder werden diese Problemen bei den ersten analytischen Schritten in Form von falschen, unverständlichen oder nicht konkludenten Beweise verortet, indem bestimmte zufällige Muster fälschlicherweise als signifikante Korrelationen interpretiert werden. Oder indem bei bestimmten Arten von Algorithmen die Verbindung zwischen den angegebenen Daten und den daraus resultierenden Korrelation nicht nachvollziehbar ist, weil die Komplexität der Algorithmen und ihre Berechnungen undurchschaubar wie eine „Black Box“ sind. Oder auch weil schließlich die Algorithmen mit schlechten Daten operiert haben, so dass die abgeleiteten „Beweise“ ebenfalls fehlerhaft sind.

Weitere Formen der Diskriminierung erfolgen hinsichtlich der Konzeptualisierung der Welt durch den Algorithmus, indem soziale Verhältnisse wie beispielsweise Vorstellungen menschlicher Würde formalisiert werden, die sich per se schwer in einer Formel erfassen lassen (Mittelstadt et al., 2016).

Das Verständnis eines ‚Ergebnisses‘ von algorithmischen Prozessen, das diesen Betrachtungen zugrunde liegt, ist jedoch in verschiedener Hinsicht problematisch. Die Ergebnisse von algorithmischen Verfahren (die Literatur dazu zentriert sich primär auf Profiling oder Personalisierung) sind Muster, die durch induktive Verfahren identifiziert werden. Es handelt sich dabei um nicht mehr als Wahrscheinlichkeitsaussagen. Weiterhin sind die identifizierten Muster nicht selbst schon ein abschließendes Urteil oder ein Handlungsvorsatz. Muster legen lediglich eine bestimmte (menschliche) Interpretation und daraus folgende Entscheidungen nahe. Es erscheint daher verfehlt, von „machine agency“, von der Maschine als Subjekt zu sprechen, die selbst „kausale Verantwortung“ trägt (Floridi, 2012). Mittels algorithmischer Verfahren können zwar vorgelagerte, automatisierte Entscheidungen getroffen werden (zum Beispiel bezüglich der Rangordnung von Postings, die auf der persönlichen Timeline eines Menschen erscheinen). Doch diese Entscheidung ist das Resultat

einer Kombination der Absichten mehrerer Akteure, die das algorithmische Verfahren (mit)gestalten. Hierbei sind beteiligt: die Designerin oder der Designer des personalisierenden Algorithmus; die Datenwissenschaftlerin oder der Datenwissenschaftler, die oder der den Algorithmus nur mit bestimmten Daten trainiert und in seiner Entwicklung laufend mitgestaltet; aber auch der Individuen, an denen dieser personalisierende Algorithmus sich orientiert und an die er sich anpasst. All diese Akteure beeinflussen das algorithmische Verfahren. Dem Automatisierungsverfahren eine kausale Verantwortung zuzuordnen – auch bei komplexeren Algorithmen –, verkennt die Bedeutung des Geflechts zwischen dem Algorithmus und den mitgestaltenden Akteuren in seinem Kontext.

All diese Herangehensweisen setzen auf eine wie auch immer geartete Form der Kontrolle und Einschränkung des Epistemischen, also desjenigen, was mittels Automatisierung in Erfahrung gebracht werden kann. Dies geschieht nicht primär aus Privatheitserwägungen, sondern in der Annahme, dass bestimmte Erkenntnisse oder Informationen missbraucht werden könnten. Dies ist jedoch eine Vorwegnahme der Handlungsoptionen aller Akteure. Durch Automatisierung und Algorithmen nämlich könnten auf der anderen Seite auch viele unentdeckte Muster der illegitimen oder unerwünschten Ungleichheit



Von der Konzeption über die Entwicklung bis zur Anwendung von Systemen zur automatisierten Entscheidungsfindung sind verschiedene Akteure beteiligt. Sie alle haben Einfluss auf die Ergebnisse und tragen somit auch einen Teil der Verantwortung.

identifiziert werden. Dies könnte dafür genutzt werden, Diskriminierung zu erkennen und zu steuern. Letztere Möglichkeit wird mit den beschriebenen Ansätzen verhindert. Zudem wird jegliche Möglichkeit des ethischen Handelns im eigentlichen Sinne verwehrt. Ethisches Handeln entsteht nicht durch wenige, eingeschränkte Informationen. Ganz im Gegenteil: Je mehr Informationen ein Individuum hat, desto besser kann es ethisch kontextualisieren und dementsprechend gerecht handeln.

Stattdessen werden bestimmte Daten, die eine öffentliche Dimension haben, als private Daten deklariert, um Diskriminierungen vorzubeugen. Es lässt sich daran die Tendenz erkennen, persönliche oder gar – im datenschutzrechtlichen Sinne – personenbezogene Daten mit privaten Daten gleichzusetzen, um das Risiko von Diskriminierungen zu beschränken. Ethische Konflikte, die eigentlich Thema einer gesellschaftlichen Debatte sein sollten, werden somit kategorisch beiseitegeschoben.

Die Konflikte werden dadurch jedoch nicht verschwinden, sondern sich womöglich verschärfen. Ein (etwas pointiertes) Beispiel: Die Schwangerschaft einer Frau gilt als eine persönliche Angelegenheit. Jedoch lässt sich ab einem bestimmten Zeitpunkt eine Schwangerschaft nicht mehr vor der Öffentlichkeit verbergen. Eine Schwangerschaft nicht nur als persönliche, sondern als ‚rein private‘ Angelegenheit zu deklarieren (mit dem Ziel, die mögliche Diskriminierung von Schwangeren zu verhindern), würde es erforderlich machen, die Schwangere vor den Augen der Öffentlichkeit zu verbergen und die Frau für die Zeit ihrer Schwangerschaft von jeglicher sozialer Teilhabe auszuschließen. Unter einer solchen Politik wären Schwangere zudem nicht mehr „normal“ im Öffentlichen. Eine gesellschaftliche Auseinandersetzung über Schwangerschaften und Inklusion im Öffentlichen (beispielsweise im beruflichen Bereich) würde nicht mehr stattfinden. Ähnliches Szenarien könnte man sich für Homosexuelle ausmalen, für Menschen mit einer nicht-weißen Hautfarbe oder mit einer sichtbaren Krankheit. In all diesen Fällen können Konflikte nicht dadurch gelöst werden, dass Informationen privatisiert oder unterdrückt werden. Auf diese Weise wird nur jegliche Identifizierung und Aufsicht über diskriminierende Verhaltensweisen unmöglich gemacht.

Ein Zwischenfazit: Alle drei bisher diskutierten Vorschläge setzen an Einschränkungen mithilfe technologischer Prozesse an. Das gilt gleichermaßen für ‚Wächter-Algorithmen‘; für verarbeitungsbezogene Ansätze (wie Zweckbindung) und datenerhebungsbezogene Ansätze (wie das Verbot, bestimmte Datenkategorien miteinander zu korrelieren). Dies geschieht wohl in der Annahme, die Orientierung am technologischen Prozess böte hinreichende Abstraktheit und zugleich Kontextualisierung, um lang anhaltende ethische Prinzipien und Einschränkungen zu formulieren. Doch diese Annahme verkennt, dass gerade digitale Technologien in ihren Grundsätzen unter permanenten Änderungen stehen – während soziale Konflikte sich in ihren Grundsatz kaum im Laufe der Geschichte geändert haben. Ethische Prinzipien, die technikorientiert sind, können per se nicht ethisch sein. Sie fokussieren lediglich auf technische Prozesse, um unethisches Handeln zu verhindern. Menschliches Handeln kommt in diesen Prozessen überhaupt nicht vor. Gerade dies jedoch ist Gegenstand der Ethik im eigentlichen Sinne: das (gute) menschliche Handeln.

In den folgenden Absätzen zeigen wir, dass es sich hierbei nicht lediglich um eine begriffliche Spitzfindigkeit oder eine Frage der Definition handelt, sondern dass Algorithmen tatsächlich bereits auf technischer Ebene fundamentale Fähigkeiten abgeben, die für verantwortliches Handeln notwendig vorausgesetzt werden müssen.

Über die Kodierbarkeit von Entscheidungen und Ethik

Bevor die Frage nach angemessenen ethischen Kategorien in der Automatisierung beantwortet werden kann, sind die Begriffe Entscheidung und Verantwortung zu definieren. In einem zweiten Schritt ist zu überprüfen, ob in algorithmischen Prozessen ethische Kriterien kodiert werden können.

Menschliches Entscheiden und Verantwortung

Menschliches Handeln (oder Handeln im Allgemeinen) gründet auf Intentionen. Verhaltensweisen, die durch Absichten (Intentionalität) kontrolliert und motiviert sind, gelten als Handlung.

Bei der Frage der Intentionalität ist wiederum zwischen Entscheidungen und Motivationen zu unterscheiden (Nida-Rümelin, 2013; Anscombe, 1963; Bratman, 1987 und 1991). Motivationen begründen sich durch Erwartungen in Bezug auf bestimmte Konsequenzen, nicht durch eine bestimmte Art des Handelns. Entscheidungen hingegen gelten bereits durch die entsprechende Handlungen als ausgeführt, unabhängig davon, ob die daraus zu erwartenden Konsequenzen erfüllt wurden.

Handlungen sind Verhaltensweisen, bei denen der Handelnde also eine Verantwortung für die Kontrolle seiner Absichten trägt. Die Kontrolle der Absichten wird durch Gründe geleitet. Ein Mensch hat letztlich nur dann Kontrolle über seine Handlungen, wenn er diese begründen kann. Gründe, Überzeugungen und Handlungen eines Menschen sind zusammen zu betrachten und auf Kohärenz zu prüfen, um Verantwortung zuordnen zu können. Die Kohärenz zwischen Überzeugungen, Gründen und Handlungen ist eine Form der Rationalität struktureller Art. Es geht nicht um den rationalen Inhalt der Überzeugungen, sondern um die prozedurale Rationalität der Handlung. Ob dieser Mensch also Verantwortung trägt, hängt vom Grad seiner Rationalität ab – aber auch vom Grad seiner Handlungsfreiheit. Verantwortung ist folglich ein graduelles Konzept: Der handelnde Mensch kann mehr oder weniger Verantwortung tragen, abhängig von

seinem Zustand und dem Kontext beim Treffen einer Entscheidung. Die emotionale Lage, der Gesundheitszustand, das Alter (Minderjährig- vs. Volljährigkeit), die Handlungsoptionen (wie viel Freiheit und Handlungsalternativen hat der Handelnde?) sind Faktoren, die für die Zuschreibung von mehr oder weniger Verantwortung sowohl ethisch als auch rechtlich eine Rolle spielen. Somit ist ein autonom handelnder Mensch niemals ein absolut autonomes Wesen, sondern er steht in Relation zum Sachverhalt und zum gesellschaftlichen Kontext und ist damit – jedenfalls in der äußeren Wahrnehmung und nach ethischen Maßstäben – abhängig von diesen.

So ist die Freiheit des Individuums, die auf seinen (rationalen, emotionalen) Absichten aufbaut, durch (gute oder viele) Gründe zu untermauern (Saake, Nassehi 2004). Dementsprechend zu entscheiden ist Ausdruck der Willensfreiheit eines Menschen. Und doch ist auch dann, wenn sich das Individuum noch so sehr bemüht, seine Entscheidung zu begründen, die Willensfreiheit nicht absolut. Der freie Wille ist nicht vorsozial. Er ist kontextuell und versöhnt – so Hegel – Notwendigkeit mit Einsicht. Armin Nassehi (2011) zufolge ist auch Zeit ein konstitutives Element dieses Willens.

Der freie Wille ist ein sozial formierter Wille. [...] Man soll nicht alles wollen, sondern alles wollen, was der eigenen kulturellen Vorstellung der menschlichen Natur entspricht. Algorithmen können den sozialen Willen mitgestalten und nachhaltig ändern. Algorithmen tun so, als wäre diese Vorstellung kontextunabhängig. Zwar kann man annehmen, dass die menschliche Natur, das Anthropologische, unbeweglich ist, von der Zeit nicht abrückbar, dennoch ist der Wille nachweislich faktisch – aber auch ethisch – im Fluss (Nassehi, 2011, 260)

Wenn man nach der Verantwortung von Maschinen oder Algorithmen fragt und über ethische Kriterien für Algorithmen nachdenkt, ist dem zu Folge die Frage der Rationalität und der (Handlungs)Freiheit bei Maschinen zu untersuchen.

Rationalität und (Handlungs-)Freiheit unter Bedingungen der Automatisierung

Algorithmen sind Mechanismen, die zu einer bestimmten Art von Resultaten führen. Die Prozesse, die sich innerhalb von komplexeren Algorithmen abspielen, sind kausaler Natur. Handeln jedoch baut auf Gründen und auf der Freiheit auf, aus verschiedenen Optionen die Handlungsalternative zu wählen, die erwünschten Entscheidungen oder Motivationen entsprechen. Die Zuschreibung von Verantwortung für Handlungen geschieht immer auch in Bezug auf diese Gründe und Freiheit. Rationale Begründungen sind logische Prozesse. Einigermaßen komplexe logische Prozesse jedoch können durch algorithmische Prozesse nicht formalisiert werden. Dies gilt in Philosophie und Wissenschaftstheorie seit Arbeiten von Church und Turing aus den 1930er Jahren als unwiderlegte Tatsache: Algorithmische Prozesse können beispielsweise die Wahrheit einer Formel – einer der elementarsten Teile der Logik – wie „Sokrates ist ein Mensch“ nicht nachweisen. „Sokrates ist ein Mensch“ ist eine prädikatenlogische Formel erster Ordnung. Die Prädikatenlogik erster Ordnung ist ein Zweig der mathematischen Logik. Die Prädikatenlogik befasst sich mit der Formalisierung von Argumenten und der Überprüfung ihrer Gültigkeit. Sie ist für Disziplinen wie Informatik, Mathematik, Linguistik und Philosophie von großer Bedeutung. Die Prädikatenlogik erster Ordnung beschäftigt sich mit den logischen Schlussfolgerungen über gewisse mathematische Ausdrücke. Dieses Schlussfolgern geschieht rein syntaktisch, das heißt ohne Bezug zu mathematischer Bedeutung. Die Prädikatenlogik erster Stufe hat zu wichtigen Erkenntnissen für die gesamte Mathematik, aber auch in der Philosophie geführt.

Prädikatenlogische Formeln können in einigen Welten wahr sein, in anderen falsch. In den 1930er Jahren haben Church und kurz danach Turing nachgewiesen, dass algorithmische Prozesse wie die Turing-Maschine Theoreme der Prädikatenlogik erster Ordnung auf deren Gültigkeit nicht überprüfen können. Bisher sind Church und Turing nicht widerlegt worden. Die Turing-Maschine kann weder beweisen, dass „Sokrates ist ein Mensch“ wahr ist, da der Satz nicht allgemeingültig ist, noch kann sie beweisen, dass „Sokrates ist kein Mensch“ wahr ist, da der Satz ebenfalls nicht allgemeingültig ist. Wenn Logik als wichtiger Bestandteil des Begründens betrachtet wird, dann ist Begründen nicht algorithmisch.

Kausale Relationen („Wenn ich genaues Wissen über einen Sachverhalt und alle relevanten Regeln darin habe, kann ich den nachfolgenden Sachverhalt determinieren“) sind zwar algorithmisch, aber das Begründen ist kein kausa-

ler, sondern ein logischer Prozess, bei dem nicht allein die Ursachen, sondern der Inhalt von Begründungen oder Argumenten zählt und nicht nur deren formales Vorhandensein.

If [...] we accept a certain understanding of causal relations, which claims that causal relations are algorithmic (i.e. that if I have exact knowledge of the current state of affairs and all relevant laws, I can determine the next state of affairs), then it has been proven since the 1930s that reasoning, in which logical inferences play a role that amount to the complexity of first-order predicate logic, is not a causal process (Nida-Rümelin, 2014)

Wenn Maschinen oder algorithmische Prozesse keine komplexen logischen Operationen vollziehen können, dann gilt erst recht, dass sie auch keine ethischen Überlegungen anstellen können. Es mangelt ihnen also mithin an jener Form von Rationalität, die für die Zuschreibung von Verantwortung maßgeblich ist. Und nicht nur das: Es mangelt ihnen auch an Freiheit. Algorithmische Prozesse besitzen keine Fähigkeit, autonom Entscheidungen zu treffen. Denn autonome, selbstbestimmte Handelnde, sind nur solche Wesen, die Handlungsfreiheit besitzen. Diese Freiheit findet ihren Ausdruck in dem moralischen Handeln, also in jener Handlungsweise, die sich begründen lässt und, frei nach Aristoteles, ihren Ursprung im Handelnden selbst hat, unabhängig von Dritten oder der äußeren Umgebung – und demnach von unmoralischen Leidenschaften unabhängig ist, da diese relationaler Natur sind (Neid, Wut etc.). Daraus folgt, dass dem maschinellen Prozess an sich keine Verantwortung zugeordnet werden kann.

So kann Verantwortung nur dem im algorithmischen Prozess involvierten und lenkenden Mensch (siehe Abbildung S. 7) zugeordnet werden. Aber auch Verantwortung wird in ihrer Gradualität von dem, was man wissen kann – den Handlungsoptionen – und dem, was man kontrollieren kann bedingt. Komplexere algorithmische Prozesse, bei denen die involvierten Menschen die weitere Entwicklung des Prozesses nicht voraussehen und nur bedingt kontrollieren und mitgestalten können, werden derzeit als Situationen mit potentiellen Risiken betrachtet und präventiv normiert, häufig mit grundsätzlichen Verboten. In der Disziplin der Entscheidungstheorie gelten obige Sachverhalte als Situationen der Ungewissheit. Diese Betrachtungsweise ermöglicht eine andere Umgangsweise: In der Entscheidungstheorie sollte bei Ungewissheit nach dem Prinzip gehandelt werden, den drohenden Höchstschaden möglichst weitgehend zu minimieren oder gar auszuschließen.

In dieser Hinsicht kann Verantwortung auch den Menschen zugeordnet werden, die in einem konkreten algorithmischen Schritt eine Möglichkeit der Mitgestaltung oder Kontrolle haben – auch wenn sie das Design und den Verlauf des Prozesses nicht im Allgemeinen verantworten (wie beispielsweise die Teams von Algorithmikern und

Datenwissenschaftlern, die bei Facebook arbeiten). Möglichkeiten der Mitgestaltung und der Kontrolle reichen uns, um eine Verantwortung zum Eingreifen zu begründen, ähnlich einer Obliegenheit im Sinne einer Garantienpflicht (wie etwa die Pflicht zur Hilfeleistung bei Unglücksfällen).

Ethik im Raum: Eine Taxonomie des Gesellschaftlichen im automatisierten Zeitalter

Prozesse, in denen automatisiert Daten verarbeitet werden, sind regulatorisch zuerst in personenbezogene und nicht personenbezogene Prozesse eingeordnet und anschließend nach Anwendungsgebieten in Subkategorien klassifiziert (Gesundheit, Finanzen, Verkehr etc.). Verwenden diese Prozesse personenbezogene oder personenbeziehbare Daten, werden sie primär im Datenschutzrecht reglementiert.

Algorithmische Prozesse, die Datenbanken ohne personenbezogene Daten verwenden, können von ebenso hoher gesellschaftlicher Relevanz sein. Sie können menschliche Kollektive mittelbar oder unmittelbar steuern. Es ist daher notwendig, eine alternative Kategorisierung vorzunehmen, die verschiedene Arten der unmittelbaren Betroffenheit von Menschen durch algorithmische Prozesse angemessener strukturiert und sich nicht an der technikorientierten Ebene der Datenverarbeitung, sondern am sozialen Kontext der Anwendung orientiert.

Bei der Erarbeitung ethischer Kriterien in Bezug auf Automatisierungsprozesse fehlt zudem eine grundsätzliche Unterscheidung zwischen Kriterien für Kollektive und für Individuen. Ethische Kriterien für Kollektive folgen einer anderen Art der Legitimation als im Falle von Sachverhalten, die Individuen betreffen. So verhält sich beispielsweise die Frage des Eigentums eines Individuums in der ethischen Abwägung anders als die Frage des Eigentums einer Gemeinschaft, bei der individuelle Ansprüche zugunsten des Kollektivs zurück treten können. Auch – wie sich am Beispiel der Allmendeproblematik verdeutlichen lässt – können Individuen mit ethischen Kriterien, die sich auf das Verhalten des Individuums konzentrieren, ethische Probleme auf kollektive Ebene nicht lösen. Ein weiterer Aspekt: Was das Individuum betrifft, so werden ethische Kriterien oft in Bezug auf Grundrechte (wie Privatheit) definiert. Ganz anders verhält es sich bei den Kollektivgütern. Vergleichbare Grundrechte, auf die sich ethische Kriterien beziehen könnten, gibt es hier nicht. Die Notwendigkeit, kollektive Güter zu schützen, bedarf also eines anderen Begründungsrahmens.

Deshalb schlagen wir auf der ersten Ebene eine Kategorisierung vor, die diese Unterscheidung aufgreift. Die Kategorie, in der algorithmische Verfahren sich an das Kollektiv richten, nennen wir die Kategorie der gesellschaftlichen Güter. Die algorithmischen Verfahren, die sich dem Individuum widmen, nennen wir die Kategorie der Individualgüter.

Man könnte in Bezug auf diese Kategorisierung entgegennehmen, dass eine große Mehrheit der Algorithmen, die von Individuen benutzt werden, personalisierende Algorithmen sind. Diese Algorithmen interessieren sich nicht für konkrete Individuen an sich, sondern für Muster, die diese Individuen in verschiedene Arten von Profilen einordnen. Daher wäre es nicht abwegig, diese ebenfalls in die Kategorie der sozialen Güter aufzunehmen – und somit den Sinn dieser Kategorisierung zu hinterfragen. Jedoch reden wir hier von zwei grundsätzlich unterschiedlichen Logiken des Handelns.

Theorien wie die „Wir-Rationalität“ (Smerilli, 2008) machen plausibel, dass individuelle Entscheidungen zum Wohl des Kollektivs weder in Bezug auf die rationale Begründung noch in Bezug auf die psychologische Motivation von verdecktem Eigeninteresse geleitet sein müssen, sondern tatsächlich auf des Gemeinwohl ausgerichtet sein können. Der Theorie der „Wir-Rationalität“ nach gibt es zwei Modi der Rationalität, nach denen Menschen handeln: die „Wir-Rationalität“, die sich am Gemeinwohl und Kollektiv orientiert, und die „Ich-Rationalität“, die sich auf die individuellen, eigenen Interessen konzentriert. Beide Modi können sich gegenseitig ausschließen. Diese Theorie erklärt, warum Individuen, die gegen das eigene Interesse handeln, nicht irrational agieren, sondern eine andere Ebene der Interessen und Absichten bei ihrem Handeln berücksichtigen.

Individuen gebrauchen also unterschiedliche Rationalitäten für individuelle und für kollektive Interessen. Eine Ethik, die sich auf Kollektivgüter und auf Kollektive richtet, folgt einer anderen Logik als die Logik individualistischer Interessen. Wie im Fall des Predictive Policing reicht eine

individualistische Ethik nicht, um die gesellschaftliche Schiefelage im Beispiel zu nivellieren.

Aussichtsreich erscheint daher auch die Berücksichtigung von Entscheidungstheorien der Wir-Rationalität für die Ausarbeitung von ethischen Kriterien, die den kollektivistischen Aspekten von algorithmischen Mechanismen Rechnung tragen.

Die bisherige ethische (und rechtliche) Debatte, die im obigen Literaturüberblick kritisch zusammengefasst wurde, fokussiert sich auf Sachverhalte, die Individualrechte betreffen. Demgegenüber werden Belange des Allgemeinwohls selten in die ethischen Überlegungen zu Algorithmen und künstlicher Intelligenz einbezogen. Dabei sind algorithmische Verfahren primär von einer kollektivistischen Herangehensweise geprägt. Wie im Fall der Frage der Diskriminierung, sind ethische Konflikte in algorithmischen Verfahren Konflikte kollektiver Natur: eine Diskriminierung erfolgt zwar individuell, aber nicht ad personam. Die Zuordnung zu einem Kollektiv ist dabei Grund und Grundlage. So wird Herr M. nicht als Herr M. von Neonazis belästigt, sondern aufgrund seiner dunkleren Haut und durch die Zuordnung zum Kollektiv etwa der „Flüchtlinge“. Paradoxiertweise ist dabei die soziale Konstruktion von Kollektiven sowohl Auslöser von Diskriminierung, als auch Referenzparameter, um Diskriminierungsmuster zu überprüfen. Die Moderne antwortet auf Diskriminierung mit dem Zugeständnis von Individualrechten aller Bürger. Jedoch reichen Individualrechte nicht, um bestimmte Kollektive strukturell zu schützen. Auch viele der Probleme, die sich als eine Form der kollektiven Diskriminierung ergeben, sind individualrechtlich nicht adressierbar. So gesehen kann Predictive Policing mit dem Einsatz von Algorithmen, die keine personenbezogenen Daten verarbeiten, eine ganze Stadt in ein soziales Ungleichgewicht versetzen, indem diese Algorithmen de facto zu sogenannten No-Go-Areas beitragen, weil eine übermäßige Polizeipräsenz statt zusätzlicher Sicherheit ein massives Sicherheitsproblem suggerieren kann. Hierbei werden weder Individuen diskriminiert (diese können ja innerhalb der Stadt umziehen) noch sind sie datenschutzrechtlich betroffen. Das heißt: Ethik und Recht, die Kollektive als Gruppen und deren Logiken nicht in den Fokus nehmen, weisen blinde Flecken auf, die einen hohen Anteil der Probleme und Risiken verbergen, die an automatisierten Verfahren haften. Aus diesem Grund erscheint eine Erweiterung der bisher verwendeten Kategorien notwendig, um Sachverhalt und Verantwortung im algorithmischen Verfahren adäquater zuzuordnen.

Gesellschaftliche Güter

Gesellschaftliche Güter beziehen sich stets auf Kollektive. Unabhängig für die Entstehung eines Kollektivs ist das Vorhandensein von verschiedenen gesellschaftlichen Rahmen, die sowohl Größe, als auch Formen der Interaktion dieses Kollektivs bestimmen. Diese gesellschaftlichen Rahmen können Inklusion oder Spaltung in einer Gesellschaft erzeugen. Innerhalb dieser Rahmen können Individuen ihre Grundrechte ausüben. Die gesellschaftlichen Rahmen bilden auch den Zugang zu Kollektivgütern.

Die Kategorie „gesellschaftliche Güter“ erfasst also algorithmische Verfahren, die sich mit Interaktionen von Kollektiven (mittels derer ein gesellschaftlicher Rahmen etabliert wird), aber auch mit Kollektivgütern befassen. Diese Kategorie unterteilen wir daher in die Unterkategorien

- gesellschaftlicher Rahmen und
- Kollektivgüter

Diese Unterkategorien sind notwendig, da hinsichtlich der Verantwortungszuordnung ethisch relevante Unterschiede zwischen den beiden festzustellen sind.

a) Der gesellschaftliche Rahmen

Der gesellschaftliche Rahmen ist die Tür zur öffentlichen Interaktion, zur Ausübung von bestimmten individuellen Grundrechten (beispielsweise Meinungs- und Versammlungsfreiheit) und für den Zugang zu den Kollektivgütern. Der gesellschaftliche Rahmen kann verschiedene Formen und Ausprägungen haben: Es ist der Hauptplatz eines Dorfes, in dem Menschen samstags ihr Gemüse einkaufen, Kinder spielen, Freunde sich verabreden. Es kann auch eine soziale Plattform wie Facebook oder Twitter sein, auf der Partys organisiert, Fanseiten eingerichtet oder Gruppen von Twitter-Nutzern in Listen aggregiert werden. Auch eine Suchmaschine wie Google, in der Informationen unter anderem nach personalisierter Relevanz in Rankings arrangiert werden, gehört dazu. Plattformen nutzen algorithmische Verfahren, die diese Interaktionen regulierend gestalten.

Der gesellschaftliche Rahmen ist eine Art Infrastruktur, die den Zugang einer Gesellschaft zum Öffentlichen und zu Kollektivgütern ermöglicht und reguliert: Eine Stadt ohne Bürgersteige, wie in vielen Orten Amerikas, bietet einen anderen Zugang, als eine Stadt wie Amsterdam mit ihren Geh-, Auto- und Fahrradwegen. Verkehrsnetze mit ihren Verkehrsregeln und Kommunikationsnetze sind ebenfalls von diesem Konzept erfasst. Aber auch digitale Plattformen konstituieren eine Form von gesellschaftlichem Rah-

men: Sie sind der neue digitale Marktplatz. Mittels ihrer Formate und algorithmischen Verfahren ermöglichen sie Kommunikation und machen Kollektivgüter wie Wissen zugänglich. (Nebenbei: Auch die Art und Weise, wie staatliche Sicherheit strukturiert und vollzogen wird, ist Teil des gesellschaftlichen Rahmens.)

Der gesellschaftliche Rahmen hat zwei Hauptcharakteristiken, die eine Unterscheidung zwischen Kollektivgütern und gesellschaftlichen Rahmen notwendig machen:

- **Kontrolle und Gestaltbarkeit:** Die Kontrolle und Gestaltbarkeit des gesellschaftlichen Rahmens obliegt einer beschränkten Gruppe von Individuen. Dies gilt sowohl für die Architektur der Stadt, als auch für die Verkehrs- und Kommunikationsnetze, die Such- und Timeline-Algorithmen, sowie für die Sicherheit. Es sind Aufgaben, die nicht kollektiv gestaltet und kontrolliert werden. Sie stehen vielmehr außerhalb einer solchen offenen Form der Gestaltbarkeit – gerade bei den gesellschaftlichen Rahmen (Architektur der Stadt, Sicherheit), die von der öffentlichen Hand zur Verfügung gestellt werden, um die Teilhabe und Zugang aller zu ermöglichen. Sie unterliegen vielmehr strengen Regeln und Rechenschaftspflichten. Dies wirft nun die Frage auf, ob gesellschaftliche Rahmen, die nicht von der öffentlichen Hand bereit gestellt werden, nicht auch besonderen Rechten und Pflichten unterliegen müssten.

- **Zugang:** Der gesellschaftliche Rahmen bestimmt, welche Teile eines Kollektivs in welcher Form Zugang zu Kollektivgütern erhalten und auf welche Art und Weise Individualrechte im Öffentlichen ausgeübt werden.

b) Kollektivgüter

Kollektivgüter sind die zweite Unterkategorie von sozialen Gütern, die von algorithmischen Verfahren erfasst werden.

Kollektivgüter können, ähnlich wie der Begriff des Allgemeinwohls oder auch der Gerechtigkeit, nicht inhaltlich definiert werden. Viel zu sehr hängen sie vom Kontext einer Gesellschaft ab. Sie werden durch prozedurale Bedingungen oder negativ (das, was sie nicht sind) beschrieben.

Von einem normativen Standpunkt müssen Kollektivgüter jedem zugänglich, von jedem nutzbar und vom Kollektiv gestaltbar sein. Die hier beschriebenen Faktoren unterscheiden sich essentiell zu den oben in Bezug auf den gesellschaftlichen Rahmen beschriebenen Faktoren. Daraus ergeben sich grundsätzliche Konsequenzen hinsichtlich der Pflichten und Verantwortungszuordnung: Der gesellschaftliche Rahmen wird von einer beschränkten, klar identifizierbaren Anzahl von Akteuren verantwortet.

Die Verantwortungszuordnung bei Kollektivgütern hingegen ist komplizierter (beispielsweise bei algorithmischen Open-Source-Verfahren, die von der Gemeinschaft weiterentwickelt werden). Ferner haben die Individuen des Kollektivs nicht nur ein Recht zur Nutzung des Gutes; dieses Recht geht Hand in Hand mit der Pflicht, zum Fortbestand des Kollektivgutes beizutragen.

Auch hinsichtlich der Ersetzbarkeit verhalten sich Kollektivgüter anders als gesellschaftliche Rahmen. Gewisse Kollektivgüter, wie etwa Wissen, sind nicht ersetzbar. Jedes Stück Information, das gelöscht wird, kann nicht durch alternative Informationen ersetzt werden. Im Gegensatz dazu kann ein gesellschaftlicher Rahmen obsolet und durch eine neue Art des Zugangs zum Öffentlichen ersetzt werden. So ist es denkbar, dass aktuell erfolgreiche Suchmaschinen oder Social-Media-Plattformen von anderen Formaten ersetzt werden. Zugleich haben die verschiedenen gesellschaftlichen Rahmen eine gewisse Monopolstellung – oder anders ausgedrückt, Konkurrenten innerhalb ihres Formates besitzen nicht dieselbe Relevanz. So hat beispielsweise Google eine gewisse Monopolstellung innerhalb seines Formates, genauso Facebook oder Twitter, und alle drei bieten unterschiedliche Zugänge und Formen der Interaktion. Sie sind nur eine Dimension, eine Form des Zugangs zum Öffentlichen: Analoge Zugänge zum Öffentlichen bieten eine Palette an Angeboten an, die nicht mit den digitalen Angeboten gleichzusetzen sind.

Ein Verzicht auf gesellschaftliche Güter impliziert, auf ein Stück sozialer Teilhabe zu verzichten.

Individualgüter

Individualgüter unterteilen wir wie folgt:

- selbst-erwählte Leistungen (wie Gesundheits-Tracker, Spiele oder Musik-Apps)
- Leistungen, die ein Individuum betreffen, aber von Drittparteien genutzt werden (Scoring oder Assistenzsysteme zur Entscheidung über Visavergabe)

Individualgüter haben einen anderen kontextuellen gesellschaftlichen Rahmen. Zumindest für einen großen Teil der algorithmischen Güter oder Dienste, die vom Individuum selbst erwählt wurden, dürfte ein Verzicht oder die Nutzung einer Alternative möglich sein. Die Möglichkeit des Verzichts lässt keine Aussage über den Grad der Sensibilität dieser Dienste zu. Individuen haben eine relative Kontrolle über die Dienste. So verlangen Konflikte, die darin entstehen, eine Abwägung zwischen den Rechten zweier Subjekte.

Algorithmische Dienste, die ein Individuum betreffen, aber von Drittparteien genutzt werden, sind hingegen selten Dienste, denen sich die betroffenen Individuen entziehen können. Diese Art der Dienste können von der öffentlichen Hand angewendet werden (beispielsweise als Assistenz in der Entscheidung zur Bewilligung von Sozialhilfe,

oder zur Assistenz einer Botschaft bei der Visumsvergabe), aber auch von privaten Akteuren (Kreditscoring). Über die Existenz mancher dieser Dienste ist sich das Individuum nicht einmal bewusst. Somit haben die betroffenen Individuen kaum Kontrolle über diese Art der algorithmischen Dienste – im Gegensatz zu selbst-erwählten Diensten.

Ethik und struktureller Kontext. Eine Zusammenfassung

Die gegenwärtige ethische Debatte und die daran anknüpfenden rechtlichen Forderungen fußen auf der modernen Vorstellung von Demokratie und Individualrechten. Algorithmen zeigen jedoch, dass beispielsweise Formen der Diskriminierung möglich sind, in der die Rechte der Individuen nicht betroffen sind, wie das oben angegebene Beispiel zu Predictive Policing zeigt. Vielmehr wird die Diskriminierung erst beim Vergleich zwischen verschiedenen Kollektiven sichtbar. Die Chancen und Ungewissheiten, die die Automatisierung begleiten, erschöpfen sich nicht einzig in der Frage der Diskriminierung. Es werden weitere Fragen aufgeworfen, die alle Arten von Grundrechten betreffen (Freiheits-, Gleichheits- und Teilhaberechte). Jedoch steckt ein grundsätzlicher kollektivistischer Charakter in Algorithmen, dem derzeit weder ethisch noch rechtlich Rechnung getragen wird und der auf alle diese Grundrechte Einfluss hat. Teilweise werden ethische Lücken übersehen, teilweise werden Sachverhalte falsch eingeordnet und folglich Verantwortung dadurch falsch bzw. zweifelhaft verortet, indem eine individualrechtliche Logik angewendet wird.

Die vorgenommene Taxonomie soll als Struktur dienen, um ethische Prinzipien für algorithmische Verfahren zu entwickeln, die neben dem individuellen auch den kollektiven Charakter von Automatisierungsmechanismen erfassen und verschiedene Logiken integrieren (Gemeinwohl versus Individualrechte), sodass eine komplexere und vollständigere Palette unterschiedlicher Machtasymmetrien und Missbrauchsrisiken dargestellt wird: Handelt es sich um Konflikte und Risiken, die das Allgemeinwohl oder Kollektive betreffen? Oder betreffen sie eher den öffentlichen Rahmen, in dem Kollektive interagieren und die den Zugang zu Kollektivgütern zur Verfügung stellen? Wird dadurch ein illegitimes Machtgefälle zwischen Staat und Bürger geschaffen? Oder geht es um eine Abwägung zwischen den Interessen zweier nicht-staatlicher Akteure? Welche Aufgaben übernehmen diese Akteure? Gibt es Aufgaben des Allgemeinwohls, die nicht automatisiert werden können? Sind diese Aufgaben gesellschaftsrelevant?

Derartige Fragen liegen auch den rechtlichen Grundsatzkategorien eines Rechtsstaates zugrunde. Traditionell trennen rechtstaatliche Gesellschaften zwischen Zivilrecht und Öffentlichem Recht. Die ethischen Kriterien, die für das Verhältnis Staat-Bürger gelten, können kaum aus verfassungsrechtlichen und demokratischen Gründen auch für das Verhältnis unter Privaten gelten; sei es in Bezug auf das Verhältnis Bürger-Bürger im engeren Sinne, oder zwischen Bürger und Unternehmen, die als Organisation von Privaten ebenfalls Grundrechtsträger sind – auch wenn bestimmten Arten von privaten Unternehmen besondere Rechten und Pflichten obliegen (sollten). Interaktionen zwischen Staat und Bürger folgen regulatorisch einer anderen Rechtsdogmatik als die Interaktionen zwischen Bürgern. So gelten beispielsweise die Grundrechte des Bürgers unmittelbar als eine Form von Abwehrrecht gegen den Staat, um seine Macht gegenüber dem Bürger einzudämmen. Im zivilrechtlichen Bereich hingegen, in der Recht die Interaktion zwischen Bürgern reguliert, gelten diese Grundrechte nur mittelbar, da sich stets kollidierende Grundrechte der verschiedenen Bürger gegenüberstehen, die es miteinander abzuwägen gilt – sofern der Staat überhaupt aufgefordert ist, im gesellschaftlichen Rahmen einer Schutzpflicht zugunsten eines Grundrechts hoheitlich zu handeln.

Die Einordnung hinter unseren Kategorien erfasst zweierlei. Zum einen geht es um eine grundsätzliche Trennung zwischen

a) ethischen Logiken und

b) Strukturen

des Gesamtgesellschaftlichen (gesellschaftliche Güter) und

a) individuellen Logiken und

b) Strukturen,

die also Individualgüter betreffen.

Zum anderen möchten wir eine Taxonomie anbieten, die mit der oben aufgeführten rechtliche Trennung kompatibel ist und die genutzt werden kann, um Regulierungsbedarf abzuschätzen und gegebenenfalls zu ergänzen. So ist die Unterkategorisierung der Individualgüter in selbstgewählte Dienste und Dienste, die von Drittparteien genutzt werden, in verschiedenen rechtlichen Korpora bereits vorhanden (E-Commerce-Richtlinie, Datenschutzverordnung etc.). Selbstgewählte Dienste werden weitgehend mit dem rechtlichen Prinzip der Zustimmung reguliert, während Dienste, die von Drittparteien genutzt werden, von Instrumenten wie der AGB-Kontrolle (bei nichtstaatlichen Akteuren) oder von konkreten Rechtsgrundlagen (sowohl bei staatlichen, als auch bei nichtstaatlichen Akteuren) geregelt werden.

Die Kategorie der gesellschaftlichen Güter andererseits eröffnet teilweise eine neue Dimension. Akteure, die den Rahmen, in dem das Öffentliche stattfindet, bereitstellen und damit faktisch (den gesellschaftlichen Rahmen) regulieren, haben eine besondere Macht und Verantwortung. Dieser Rahmen war bisher primär von staatlichen Akteuren reguliert, sowohl durch Institutionen und Akteure wie die Polizei oder die Grenzkontrollen, als auch rechtlich, beispielsweise durch das Demonstrationsrecht, das Äußerungsrecht etc. Presse und Rundfunk mit ihrer Funktion der Meinungsbildung und Informationsvermittlung waren bisher einige der wenigen nicht-staatlichen Akteure, die ebenfalls eine große Rolle bei der Bildung des Allgemeinöffentlichen sowie weiterer Fachöffentlichkeiten spielten und gesellschaftlichen Angelegenheiten Sichtbarkeit verleihen oder entziehen konnten. Jedoch war diese Funktion vom staatlichen Rechtsrahmen strukturiert: Der Presse und dem Rundfunk wurden aufgrund ihrer wichtigen Funktion besonderen Privilegien eingeräumt, aber auch besondere Pflichten auferlegt; der Rundfunk war sogar lange Zeit rein staatlich im Sinne eines staatlich vermittelten Pluralismus geregelt. Durch die Digitalisierung gesellen sich neue Akteure zur Dimension des Öffentlichen, die weder die Rolle der Presse, noch einer staatlichen Funktion übernehmen.

Für die Notwendigkeit der Kategorie der gesellschaftlichen Güter spricht die fehlende Berücksichtigung weiterer Aspekte des Kontextes, insbesondere der verschiedenen Dimensionen, in denen der Gesellschaft Schäden oder Nachteile zugefügt werden können. Denn diese Schäden betreffen nicht immer Individuen, wie im Falle von Predictive Policing geschildert. Sie können entweder eine Auswirkung auf einzelne Kollektive oder einen strukturellen Einfluss auf das Gesamtgesellschaftliche dadurch haben, dass sie die Unterkategorien Kollektivgüter oder den gesellschaftlichen Rahmen (den Rahmen des Öffentlichen) erfassen.

Auch die Regulierung des öffentlichen und privaten Raums wird in der Kontextualisierung nur partiell unter Revision gestellt, zumal die Digitalisierung sowie die Vorstellung des Öffentlichen und Privaten grundsätzlich von räumlichen Vorstellungen entkoppelt ist. Als Basis der rechtlichen und auch ethischen Debatte galt lange – und gilt zum Teil noch heute in einer fortentwickelten Form durch die Rechtsprechung – die Sphärentheorie, die das Private und das Öffentliche als grundsätzlich abtrennbare Räume betrachtet. Da die Vorstellung des Räumlichen durch das Internet aufgehoben wird, konzentrieren sich die Diskussionen auf Formen der Trennung dieser sogenannten Sphären. Dabei wird übersehen, dass das Kriterium des Räumlichen, um Privates vom Öffentlichen zu unterscheiden, ein Kriterium formaler Natur war, und keineswegs die gesellschaftlichen Vorstellungen des Privaten und Öffentlichen in ihrer Komplexität darstellte. Durch diese Fokussierung wurde insbesondere eine nähere Betrachtung der Strukturen des Öffentlichen beiseite gelassen. Fragen nach der Bereitstellung des Zugangs und der Strukturierung des Öffentlichen wurden nur in verschiedenen Teilbereichen (Verkehr, Presse etc.) reguliert, ohne das Öffentliche grundsätzlich zu betrachten; die Fragen des gesellschaftlichen Kollektivs und seiner Funktion im Öffentlichen wurden nur partiell behandelt, etwa in der Diskussion über Religion und das Demonstrationsrecht.

Durch Automatisierung stehen aber gerade diese Aspekte im Mittelpunkt und bedürfen der ethischen und teilweise rechtlichen Betrachtung. Die vorgestellte Taxonomie setzt an dieser Stelle an und strukturiert das Öffentliche in Kategorien, die sich nicht nur in Meinungen und Informationen erschöpfen, sondern einerseits Kollektivgüter erfassen und andererseits individuelle und kollektive Interaktionen betrachten. Darüber hinaus berücksichtigen sie die Strukturen, Zugänge und Moderation des Öffentlichen. Somit wird eine bessere ethische Kontextualisierung für die Ausarbeitung differenzierter ethischer Kriterien angeboten, die eine technikneutrale, an Werten orientierte Betrachtung wieder in Mittelpunkt stellt.

Literaturverzeichnis

- Anderson, M., Anderson, S.L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4), 15.
- Anscombe, G. E. M. (1963). *Intention*, second edition, Oxford: Blackwell.
- Arendt, H. (2002). *Vita activa oder Vom tätigen Leben*, third edition, Munich: Piper, 1967.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.
- Bratman, M. (1991). *Cognitivism about Practical Reason, reprinted in Faces of Intention*, Cambridge: Cambridge University Press, 1999.
- Bello, P., Bringsjord, S. (2012). *On how to build a moral machine*. *Topoi*, 32(2), 251–266.
- Bozdog, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227.
- Etzioni, A., Etzioni, O. AI Assisted Ethics (2016). *Ethics and Information Technology*, 18(2), 149-156. Retrieved from <https://ssrn.com/abstract=2781702> (accessed 2 October 2016).
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology*, 25(4), 435–437.
- Friedman, B., Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), 330–347.
- Hildebrandt, M. (2011). Who needs stories if you can get the data? ISPs in the era of big number crunching. *Philosophy & Technology*, 24(4), 371–390.
- Kant, I. (1977). *Werke in zwölf Bänden*. Band 7, Frankfurt am Main.
- Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, 45(5), 494–511.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).
- Naik, G., Bhide, S. (2014). Will the future of knowledge work automation transform personalized medicine? *Applied & Translational Genomics, Inaugural Issue*, 3(3), 50–53.
- Nassehi, A. (2011). *Gesellschaft der Gegenwart: Studien zur Theorie der modernen Gesellschaft II*. Berlin: Suhrkamp Verlag.
- Nida-Rümelin, J. (2014). Agency, technology, and responsibility. *Politica & Società*, 2, 185-200.
- Nissenbaum, H. (2001). How computer systems embody values. *IEEE Computer*, 34, 118-120.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.
- Rawls, J. (1979). *Eine Theorie der Gerechtigkeit*. Frankfurt: Suhrkamp Verlag.
- Romei, A., Ruggieri, S. (2014). A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review*, 29(5), 582–638.
- Saake, I., Nassehi, A. (2004). Die Kulturalisierung der Ethik. Eine zeitdiagnostische Anwendung des Luhmannschen Kulturbegriffs. In: Günter Burkart and Gunter Runkel (eds.): *Niklas Luhmann und die Kulturtheorie*, Frankfurt/M.: Suhrkamp: 102-135.
- Schermer, B.W. (2011). The limits of privacy in automated profiling and data mining. *Computer Law & Security Review*, 27(1), 45–52.
- Smerilli, A. (2008). We-thinking and ‚double-crossing‘: frames, reasoning and equilibria, MPRA Paper, University Library of Munich. Retrieved from https://mpra.ub.uni-muenchen.de/11545/2/MPRA_paper_11545.pdf (accessed 2 October 2016).
- Tene, O., Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. Retrieved from: http://heonlinebackup.com/hol/cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11§ion=20 (accessed 2 October 2016).
- Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, 9(1), 49–62.
- Tutt, A. (2016). An FDA for algorithms. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Retrieved from: <http://papers.ssrn.com/abstract=2747994> (accessed 03 November 2016).
- Wehofsits, A. (2016). Die ethische Dimension von Big Data. Vodafone Institute. Retrieved from http://www.vodafone-institut.de/wp-content/uploads/2016/10/Big-Data_Ethische-Fragen.pdf (accessed 7 November 2016).
- Wiltshire, T. J. (2015). A prospective framework for the design of ideal artificial moral agents: Insights from the science of heroism in humans. *Minds and Machines*, 25(1), 57–71.

Autorinnen

Lorena Jaume-Palásí ist Mitgründerin von AlgorithmWatch und forscht zu Rechtsphilosophie und politischer Philosophie im digitalen Zeitalter. Dabei fokussiert sie sich auf zeitgenössische Vorstellungen und die Dynamiken digitaler Öffentlichkeit und Privatheit, insbesondere auf ethische Konflikte und Normen. Unter anderem beriet sie als eine von acht externen Experten Googles Advisory Council zum Thema „Recht auf Vergessen“. Sie ist Mitgründerin der Dynamic Coalition on Publicness des Internet Governance Forums der Vereinten Nationen (UN IGF) und leitet das Sekretariat des deutschen Internet Governance Forums (IGF-D). Lorena Jaume-Palásí ist Mitglied im Expertenbeirat der internationalen Initiative „Code Red“ gegen Massenüberwachung. Derzeit ist sie Bucerius-Fellow der ZEIT-Stiftung, Mitherausgeberin von *Digitalpolitik. Eine Einführung* (Berlin 2017), *Guidebook Internet Governance* (Bonn 2016); *Keep the Internet Free, Open and Secure* (Berlin 2013, mit W. Kleinwächter); Ko-Autorin des *German Country Report in: A Crisis of Accountability. A global analysis of the impact of the Snowden revelations* (Amsterdam 2014).

Matthias Spielkamp ist Mitgründer von AlgorithmWatch und Gründungsmitglied und Herausgeber von iRights.info (Grimme Online Award). Er hatte Lehraufträge an verschiedenen deutschen Hochschulen und war Sachverständiger des Bundestags zu künstlicher Intelligenz und Robotik, Geheimdienstkontrolle, Online-Journalismus und Urheberrecht. Derzeit ist er Bucerius-Fellow der ZEIT-Stiftung; 2015/16 war er Fellow der Stiftung Mercator und Gastwissenschaftler am Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG). Spielkamp ist Mitglied im Vorstand bei Reporter ohne Grenzen Deutschland, außerdem im Beirat des Whistleblower-Netzwerks und des Studiengangs Politik & Public Affairs der Quadriga Hochschule Berlin. Im Lenkungskreis des deutschen Internet Governance Forums (IGF-D) ist er Co-Chair für die Gruppen Wissenschaft und Zivilgesellschaft. Mitherausgeber von *Digitalpolitik. Eine Einführung* (Berlin 2017); *Guidebook Internet Governance* (Bonn 2016), *Groundbreaking Journalism* (Berlin 2014); Ko-Autor von *Arbeit 2.0* (Berlin 2009, mit V. Djordjevic et al.); *Urheberrecht im Alltag* (Bonn 2008, mit V. Djordjevic et al.); *Schreiben fürs Web* (Konstanz 2003, mit M. Wieland).

Impressum

AW AlgorithmWatch gGmbH

Bergstr. 22, D-10115 Berlin

Sitz der Gesellschaft: Oranienstr. 19a, D-10999 Berlin

Amtsgericht Berlin Charlottenburg, HRB 186522 B

Geschäftsführung: Lorena Jaume-Palasi, Matthias Spielkamp

E-Mail: ljp@algorithmwatch.org | ms@algorithmwatch.org

Lizenz

Diese Publikation ist unter der Lizenz

Creative Commons Namensnennung 3.0 Deutschland

(CC BY 3.0 Deutschland) erschienen.

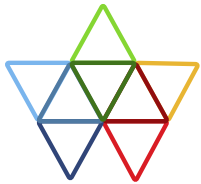


<https://creativecommons.org/licenses/by/3.0/de/>

Wir empfehlen, bei Zitaten folgende bibliografische Angaben zu machen:

Lorena Jaume-Palasi, Matthias Spielkamp: Ethik und algorithmische Prozesse zur Entscheidungsfindung oder -vorbereitung, AlgorithmWatch Arbeitspapier Nr. 4, Berlin 2017

Die Entstehung dieses Arbeitspapiers wurde durch ein Fellowship des Bucerius Labs der ZEIT-Stiftung Ebelin und Gerd Bucerius ermöglicht.



ALGORITHM WATCH

Je weiter Technik entwickelt wird, desto komplexer wird sie. Wir sind der Ansicht, dass Komplexität nicht bedeuten darf, dass Technik nicht mehr nachvollziehbar ist (siehe unser ADM-Manifest). AlgorithmWatch ist eine nicht-kommerzielle Initiative mit dem Ziel, Prozesse algorithmischer Entscheidungsfindung zu betrachten und einzuordnen, die eine gesellschaftliche Relevanz haben – die also entweder menschliche Entscheidungen vorhersagen oder vorbestimmen, oder Entscheidungen automatisiert treffen.

WIE ARBEITEN WIR?

Beobachten

AlgorithmWatch analysiert die Auswirkungen algorithmischer Entscheidungsfindungsprozesse auf menschliches Verhalten und zeigt ethische Konflikte auf.

Erläutern

AlgorithmWatch erläutert die Eigenschaften und Auswirkungen komplexer Prozesse algorithmischer Entscheidungsfindung für eine breite Öffentlichkeit.

Verknüpfen

AlgorithmWatch ist eine Plattform, die Experten verschiedener Kulturen und Disziplinen zusammen bringt, die sich mit der Analyse algorithmischer Entscheidungsfindung und ihren gesellschaftlichen Auswirkungen beschäftigen.

Mitgestalten

Um den Nutzen algorithmischer Entscheidungsfindung für das Gemeinwohl zu maximieren, unterstützt AlgorithmWatch dabei, Ideen und Strategien zu entwickeln, die eine Nachvollziehbarkeit der Prozesse ermöglichen – mit einem Mix aus Technologie, Regulierung und geeigneten Aufsichtsinstitutionen.

<https://algorithmwatch.org>

