

Algorithmenkritik und die Suche nach dem »Außen«

Tobias Matzner

Abstract: *This article reflects on critiques of algorithms in recent years, i.e. texts that critically engage with algorithms from an ethical or political perspective. It shows that many of these works derive their normativity from an »outside« of the algorithmic. For example, an algorithmic classification is criticised on the basis of a comparison with a non-algorithmic, human-made classification. Three forms of »outside« are presented. (1) The »outside« as a site of instrumental control over algorithms. Two further forms of the »outside« are found in theories that reject an instrumental understanding of technology: (2) The unpredictable, which marks an inherent limit of the algorithmic, from which some authors draw political consequences. (3) A form of social interaction that is not algorithmically (or non-technically) mediated and that is opposed to the algorithmically mediated form. This article critiques all three forms and proposes a different critique of algorithms, one that does not seek an »outside« but rather begins »within« the algorithmic itself, drawing its normativity from the difference between algorithms or algorithmically mediated situations themselves.*

Keywords: *criticism of algorithms; digital ethics; unpredictability; normativity; philosophy of technology*

In den letzten Jahren wurden eine Vielzahl von Texten veröffentlicht, die sich kritisch mit der Ethik oder Politik von Algorithmen und künstlicher Intelligenz auseinandersetzen. Dabei kommen ganz unterschiedliche Begriffe und Vorstellungen von Algorithmen – und Technologie allgemein – sowie deren ethischer oder politischer Kontextualisierung zum Tragen. In diesem Text geht es um eine argumentative Figur, die viele dieser Texte trotz der Differenzen eint: Die Kritik an einer algorithmischen Praxis oder einer algorithmischen Anwendung entsteht durch den Bezug zu einem »Außen« der Algorithmen. So wird z. B. eine algorithmische Klassifikation kritisiert basierend auf dem Vergleich mit einer nicht-algorithmischen Klassifikation, also einer Klassifikation durch Menschen. Eine ganze Reihe von Begriffsschöpfungen nach dem Muster »algorithmic X« lassen sich finden: »algorithmic gover-

nance« (Rouvroy 2013), »algorithmic cultures« (Seyfert/Roberge 2016), »algorithmic war« (Amoore 2009) und viele mehr. Oft wird hier das »algorithmic X« kritisiert, indem es auf eine nicht-algorithmische Form von X bezogen wird. Es werden jedoch nicht nur solche Differenzen zwischen algorithmischen und nicht-algorithmischen Varianten des Gleichen bemüht. Eine noch grundlegendere Form des »Außen« stammt aus Überlegungen, was Algorithmen per se nicht möglich oder zugänglich ist.

Zur Begriffsschärfung und Definition dessen, was Algorithmen sind und was sie ausmacht, ist der Bezug zu beiden Formen des »Außen« sicher richtig und wichtig. Hier geht es jedoch um etwas anderes: Die Differenz zwischen Algorithmen und dessen »Außen« wird zur Quelle der politischen Bewertung von Algorithmen.

Im Folgenden werde ich drei verschiedene Formen dieser politischen oder ethisch-politischen Argumentation über ein »Außen« der Algorithmen vorstellen. Erstens das »Außen« als Ort der Kontrolle über Algorithmen, zweitens das »Außen« als das nicht-Berechenbare oder Unberechenbare und drittens, das »Außen« als die Form sozialer Interaktion, die durch Algorithmen gerade bedroht ist. Für jeden Punkt werde ich einige Probleme anführen, die dann zum Schluss zeigen lassen, dass die hier vorgestellte Kritik von außen durch eine Kritik »innerhalb« des Algorithmischen ergänzt werden muss. Diese speist sich aus unterschiedlichen Formen und Konfigurationen von Algorithmen. Statt der Differenz zwischen Algorithmen und einem Außen betrachtet diese Kritik also die Differenz zwischen unterschiedlichen Formen oder Konfigurationen von Algorithmen.

1. Kontrolle über Algorithmen

Die erste Variante des Arguments über ein Außen speist sich aus einem liberalen oder auch instrumentellen Technikverständnis. Demzufolge wird Technik zum Problem, wenn sie nicht mehr von hinreichend autonomen Subjekten kontrolliert wird und in ihren Konsequenzen beherrschbar bleibt. Technik ist dieser Vorstellung nach ein Objekt, das einem Subjekt als Instrument für seine Ziele dienen soll. Die Politik der Technik wird also durch das Subjekt und dessen Anliegen definiert, und beides ist der Technik äußerlich gegenübergestellt. In anderen Worten kann dieselbe Technik in den Händen des einen Subjekts ein hilfreiches Werkzeug sein, in denen eines anderen ein Mittel für Missbrauch oder Gewalt. Das politische relevante »Außen« der Algorithmen sind also die Menschen, die sie nutzen, aber auch von ihrer Nutzung betroffen sind. Das instrumentelle Technikverständnis ist hier nun aber nicht deskriptiv, sondern normativ zu verstehen. Algorithmen bergen demzufolge die Gefahr, sich genau dieser Kontrolle und damit einer verantwortbaren Politik zu entziehen.

Diese Vorstellung ist sehr wirkmächtig in Ethikleitlinien und Policy-Empfehlungen. Die dort geforderten Werte wie Transparenz oder Einwilligung haben zum Ziel, den menschlichen Nutzer*innen oder Betroffenen diese Kontrolle zu sichern oder wiederzugeben. Umgekehrt speist sich Kritik dann daraus, dass diese Kontrolle über die Techniknutzung und das Verstehen ihrer Konsequenzen gerade bei Algorithmen oft nicht mehr möglich sei. Gängige Motive sind hier die sogenannte »Blackbox« der Algorithmen, womit der Zugang zu Quellcode oder Designentscheidungen gemeint ist, der lediglich Programmierer*innen möglich ist (Kitchin 2017; Matzner 2017). Im Kontext von künstlicher Intelligenz wird auch angeführt, dass nicht einmal die Entwickler*innen selbst verstehen könnten, wie die Technik wirklich funktioniert (Rohlfing et al. 2020). Ähnliche Argumente finden sich auch in Bezug auf Trainingsdaten von Machine Learning Algorithmen, deren Quelle und Verarbeitung aus Sicht der Anwendung und der Betroffenen nicht nachvollziehbar ist (Mühlhoff 2019). Eine verwandte Argumentationsweise findet sich auch in der Kritik des sogenannten Überwachungskapitalismus zum Beispiel bei Shoshanna Zuboff (Zuboff 2019). Sie argumentiert, dass aus der Tätigkeit von Menschen online digitale Produkte entwickelt werden und somit diese ohne deren Wissen als »Sklaven« arbeiteten (Zuboff 2016). In all diesen Argumenten ist die selbstbestimmte Nutzung der Technologie beziehungsweise deren Unmöglichkeit die Quelle der Kritik – noch nicht einmal die bewusste Aufgabe der Selbstbestimmung ist nach Zuboff und anderen noch möglich. Die Nutzung vieler digitaler Technologien erfolgt, ohne die Konsequenzen kennen zu können.

Diese Gegenüberstellung von menschlichem Subjekt und technischem Objekt und das ihr zugrundeliegende instrumentelle Technikverständnis werden in weiten Bereichen der Technikphilosophie, der Medienwissenschaft und der Science and Technology Studies kritisch gesehen. Hier wird betont, dass menschliche Subjektposition immer schon in einer Art und Weise durch Technik geformt ist, die von Menschen nicht komplett nachvollzogen oder durchdrungen werden kann. Zum Beispiel hat N. Kathrine Hayles in ihrer Geschichte der Kybernetik dargestellt, wie letztere einerseits als liberales Projekt verstanden werden muss, das aus dem Versuch entstand, neue technische Nutzungsmöglichkeiten zu erschließen. Andererseits hat sie aber das liberale Subjekt radikal in Frage gestellt, wenn es als komplexe Form von Informations- und Kommunikationsschleifen Tieren oder sogar Maschinen ontologisch gleichgestellt wird (Hayles 1999). In einflussreichen Teilen der Technikphilosophie wird Technik grundlegend so verstanden, dass ihr Sinn genau darin besteht, Menschen Handlungen zu ermöglichen, die sie genau nicht selbst durchführen können oder nicht mehr selbst durchführen können müssen. Das heißt, ihr Sinn besteht gerade darin, gewisse Dinge nicht zu wissen oder können zu müssen (Krämer 2021). Aus politischer Sicht lässt sich hinzufügen, dass dieses Technikverständnis einem individualistischen Politikverständnis entspricht (Matzner 2019b). Daraus folgt unter anderem, dass die Verantwortung für die Tech-

nologie reduziert wird auf die Verantwortung für die Techniknutzung durch das individuelle Subjekt. Gerade im Kontext von Algorithmen ist es aber nicht möglich, wirklich von verantwortungsvoller Nutzung zu sprechen, weil die Technik schlicht und ergreifend nicht überschaut werden kann. Selbst informatisch hervorragend gebildete Menschen können nur errahnen, was mit ihren Daten geschieht, wenn sie beispielsweise eine Webseite aufrufen. Die informierte Einwilligung welche durch Cookie-Warnungen oder das Akzeptieren von Lizenzen suggeriert wird, ist somit eigentlich nicht möglich (Matzner 2024: 147ff.). Statt diese Unmöglichkeit zu bemängeln und zu versuchen, sie dennoch herzustellen, wäre die Alternative, die Verantwortung gleich bei den Betreibern oder den Entwicklern zu suchen und nicht mehr beim Individuum selbst. Unter dem Stichwort »Responsibilisierung« wird diskutiert, wie durch solche Maßnahmen, die vordergründig so klingen, als wollten sie die Autonomie der Subjekte durch mehr Kontrolle vergrößern, diese eigentlich einschränken, indem sie sie mit einer nicht zu lösen Aufgabe überfordern (Matzner et al. 2016). Zudem stimmt die implizierte Zuschreibung, dass die Normativität einer Technik aus der Nutzung kommt, spätestens im Fall von Algorithmen in aller Deutlichkeit nicht mehr. Wenn z.B. ein Unternehmen einen Algorithmus zur Bewertung von Bewerbungen kauft, der ein Bias aufweist, dann ist das nicht dem Unternehmen oder seinen Angestellten zuzuschreiben. Das heißt, sowohl aus technisch praktischer Sicht ist es nicht möglich, eine volle Kontrolle auszuüben, wie es auch aus normativer Sicht nicht angemessen ist, diese Aufgabe den Nutzer*innen alleine zu überlassen.

Abstrakt gesprochen lassen sich diese Kritiken am instrumentalen Technikverständnis darin zusammenfassen, dass hier das Subjekt und damit die Quelle der Kritik der Technik rein äußerlich ist. Darüber hinaus ist auch das mit dieser Kritik verbundene Verständnis von Subjekten in vielen Aspekten von kritischer Theorie bis Systemtheorie kritisiert worden. Da all diese Punkte in der Literatur ausführlich diskutiert sind, belasse ich es in diesem Text mit dieser kurzen Erwähnung und gehe zur nächsten Form über, in der Algorithmen durch ein »Außen« kritisiert werden.

2. Das Unberechenbare

Eine zweite Gruppe von Theorien versammle ich hier unter dem Begriff des Unberechenbaren. Diese arbeiten nicht mit einer Gegenüberstellung von Mensch und Technik. Vielmehr finden sich hier Theorien, die der Technik sogar einen recht großen Einfluss auf menschliche Subjekte zumessen. Das Außen des Algorithmischen besteht hier viel mehr in einer prinzipiellen Unverfügbarkeit für Algorithmen. Diese ist unabhängig von deren Nutzung oder Einsatzkontext.

Genauer müssen in Bezug auf den Begriff des Unberechenbaren zwei Theorielinien unterschieden werden. Die eine bezieht sich auf Erkenntnisse der theoretischen

schen Informatik und Mathematik, in denen es um die Grenzen der Berechenbarkeit geht. Hier geht es also darum zu klären, wie Berechenbarkeit mathematisch gefasst werden kann und was in dieser Hinsicht überhaupt berechenbar ist. Wichtige Grundlagentexte der Informatik haben gezeigt, dass es tatsächlich Dinge gibt, die von Algorithmen oder, um genauer zu sein, von deren mathematischem Modell wie der sogenannten Turingmaschine oder formalen Kalkülen nicht berechenbar sind (Turing 1937; Church 1936; für eine Zusammenfassung siehe Matzner 2024: 15ff.). In dieser Tradition finden sich beispielsweise die Arbeiten von Luciana Parisi, die sich allerdings nicht direkt auf Alan Turing stützt, sondern auf den etwas obskuren Mathematiker Gregory Chaitin und seine Arbeit zum Unberechenbaren (Parisi 2016). Dennoch bezieht sich auch diese Theorie auf die grundsätzliche, formal definierte Berechenbarkeit. Hier ist im Kontext aktueller Anwendung von Algorithmen zu betonen, dass sich Berechenbarkeit in diesen Texten immer auf deterministische Berechenbarkeit bezieht, das heißt auf das garantierte Erreichen des richtigen Ergebnisses. Maschinelles Lernen ist aber genau aus dem Versuch entstanden, mit diesen Grenzen der Berechenbarkeit umzugehen, indem eben nicht deterministisch gerechnet wird. Stattdessen werden statistische, heuristische oder andere näherungsweise Rechenverfahren genutzt. Bei all diesen wird das Ergebnis nicht garantiert gefunden, sondern nur mit einer hinreichenden Wahrscheinlichkeit und hinreichender Genauigkeit erreicht. Eine Kritik, die sich daraus speist, dass Algorithmen, weil sie maschinell rechnen, gewisse Dinge per se nicht können, verfehlt also gerade Algorithmen im Sinne von maschinellem Lernen, der künstlichen Intelligenz oder vielen anderen zeitgenössischen Anwendungen, wie z.B. Optimierungsverfahren, die alle nicht deterministisch rechnen und sich somit Dingen, die Algorithmen tatsächlich deterministisch nicht können, dennoch anzunähern.

Diese erste Theorielinie des Unberechenbaren als ein Außen des Algorithmischen werde ich hier also nicht weiter verfolgen. Stattdessen beschäftige ich mich mit der zweiten Form des Unberechenbaren. Diese bezieht sich nicht auf mathematische oder maschinelle Berechenbarkeit. Stattdessen geht es um die prinzipiellen Möglichkeiten von Algorithmen in ihrem aktuellen Einsatzkontext und insbesondere um die Möglichkeiten des maschinellen Lernens. Diese Form des Unberechenbaren wurde vor allem von Louise Amoore in ihrem Buch *Cloud Ethics* ausbuchstabiert (Amoore 2020). Sie untersucht dort verschiedene algorithmische Verfahren der Prädiktion im Bereich der Sicherheit, der Genanalyse und andere mehr. Sie zeichnen sich alle genau dadurch aus, unsichere, probabilistische oder sogar die Zukunft betreffende Aussagen zu suchen. Für Amoore sind die Ausgaben von Algorithmen aber nicht nur deshalb unsicher, weil sie approximativ oder prädiktiv sind. Denn nicht einmal die im Rahmen der Statistik möglichen Garantien einer Aussage hält Amoore für gerechtfertigt.

Sie betrachtet Algorithmen nicht als Rechnung, sondern als automatisierten Akt des Schreibens. Algorithmen, egal wie sie genau mathematisch arbeiten, liefern Zei-

chen, welche eine Aussage über die Welt beinhalten. Diese Aussagen von Algorithmen sind für Amoore nun prinzipiell unsicher, weil sie so unsicher sind, wie es jeder Akt des Schreibens ist. Hier bezieht sich Amoore auf Foucaults Überlegungen zur Autorschaft (Foucault 1988) sowie auf die Theorie von Jacques Derrida zur Schrift. Derzufolge kann jede Art des Schreibens nur versuchen, seine vermeintlich sichere Bedeutung einzuholen, diese aber nie letztendlich erreichen (Derrida 2001). Algorithmen sind als Akte des Schreibens – wenn auch des maschinellen Schreibens – in diesem Sinne genauso unsicher; oder genauer gesagt, die Bedeutung des algorithmischen Schreibens kann nie abgeschlossen werden. Es ist genau diese prinzipielle Unabgeschlossenheit von Schrift, die Amoore als das Unberechenbare bezeichnet.

Damit richtet sie sich erst einmal gegen Versuche, die Ethik von Algorithmen durch Zugang oder Transparenz zu lösen:

»The calls for access to the source code of algorithms are a means to try to read how a likelihood ratio is written. Yet, scrutinizing the code itself would reveal nothing of the racialized, prejudicial outputs of the algorithm that are written via training data, validation data, and the feedback loops of experiments in the field. Indeed, the Github platform is itself a form of distributed and iterative writing in which multiple developers contribute to the rewriting and editing of software.« (Amoore 2020: 97)

Damit richtet sich Amoore auch gegen die im ersten Abschnitt diskutierte Strategie, Subjekten gegenüber der Technik Kontrolle zu geben – in diesem Fall eben durch Einsicht in den Quellcode als die entscheidende Instanz, wo Kontrolle über die Technik ausgeübt würde. Neben dem Code wird der Output von Algorithmen des maschinellen Lernens aber durch viele Faktoren bestimmt: Trainingsdaten, die Feedbackloops der Trainingsprozesse. Und auch der Code selbst ist nicht eine zentrale Instanz sondern oft verteilt und in verschiedenen Formen der Zugänglichkeit reguliert (Matzner 2024: 89ff.), wofür Amoore hier als Beispiel die Codeverwaltungsplattform GitHub angibt. All diese Faktoren sind aber nicht nur eine Quelle von großer Komplexität für Amoore. Sie stellen in vielerlei Hinsicht einen Bezug zu anderen sozialen, kulturellen, und ökonomischen Prozessen her und sind damit eine Quelle von Kontingenz. Selbst wenn also hypothetischerweise der Aufwand betrieben werden könnte, der nötig ist, um der Komplexität algorithmischer Prozesse gerecht zu werden, könnten diese nicht durchdrungen werden. Sie sind aufgrund der Kontingenzen in den Prozessen selbst kontingent und damit prinzipiell unabgeschlossen. Genau so, wie ein Text laut Foucault nicht auf einen Autor oder dessen Intentionen zurückzuführen ist, sondern seine Bedeutung durch vielerlei Aspekte erhält, die insbesondere auch in Bezügen zu anderen Texten entsteht, so lässt sich auch keine definitive Quelle der Bedeutung algorithmischer Ausgaben finden. Amoore betrachtet Algorithmen als technischen Schreibprozess, deren Bedeutung aber ebenso

relational und offen zu verstehen ist. »The algorithm iterates beyond the moment of its inscription, distributing the writing through multiple characters, from the training data to the back propagation of errors.« (Amoore 2020: 100) Damit, so Amoore weiter, sind Algorithmen gerade nichts Besonderes:

»But my point is that we should remember that algorithms are not unique, or somehow ›outside the text.‹ All acts of writing and reading, whatever the text, necessarily confront illegibility and the impossibility of reading. We do not need new resources or technologies to prize open the unreadable algorithm. Instead we could begin from that unreadability as the condition of all engagements with text. It is something familiar to us all as social and political theorists.« (Amoore 2020: 100)

Das ethische Problem besteht folglich vielmehr darin, zu denken, Algorithmen wären etwas Besonderes. Diese Besonderheit hat viele Formen: Algorithmen als eine objektive Instanz, frei von sozialen oder kulturellen Einflüssen, eine Quelle von Berechenbarkeit in der Unberechenbarkeit des Lebens, ja sogar eine Produktionsweise neuen Wissens. All diese Vorstellungen haben gemein, dass Algorithmen als etwas gesehen werden, das klare, lesbare, eindeutige Ergebnisse produziert. Für Amoore ist das Unberechenbare, das sie auch das Unlesebare oder Unattribuierbare (unattributable) nennt, also nicht das Problem, wie oben im ersten Abschnitt diskutiert. Vielmehr ist es die Quelle ihrer Ethik. Dieser geht es einerseits darum, diese Offenheit anzuerkennen, was große Konsequenzen für die Frage hat, welchen Wert algorithmischen Entscheidungen beigemessen werden. Das bedeutet für sie nicht, dass diese keinen Wert hätten. Aber sie müssen eben mit der Vorsicht und Kompetenz sozial und politisch kundiger Menschen »gelesen« werden, wie sie im obigen Zitat schreibt. Andererseits geht es auch darum, diese Offenheit zu wahren. Denn nur in ihr findet sich auch die Möglichkeit für Veränderung und Widerstand.

»In place of an ethics that seeks to make an illegible algorithm legible to the world, a cloud ethics recognizes the nonclosure in all forms of writing and pushes the fabulation of the algorithm beyond what can currently be read of its logic.« (Amoore 2020: 160)

Die hier angesprochene »fabulation«, so konstatiert Amoore, sei bereits am Werk. Denn die Informatik hätte kein Problem damit, »Fabeln« über ihre Technik zu erzählen, z.B. dass sie eine Möglichkeit maschineller Vernunft beinhalte. Auch Algorithmen selbst fabulierten: »they invent a people, write them into being as a curious body of correlated attributes, grouped into clusters derived from data that are themselves fabulatory devices« (Amoore 2020: 158). Diese als Lüge oder Ideologie zu enttarnen, verschließt sich aber aufgrund Amoores hier ausgeführtem Verständnis.

Stattdessen ist die Herausforderung einer Ethik, die Unabgeschlossenheit zu nutzen, und den technikdeterministischen oder vorurteilsbehafteten Fabulationen andere – eben ethische – beizufügen.

Dieses Betonen einer prinzipiellen Möglichkeit von Alternativen einer prinzipiellen Offenheit ist ein erster wichtiger Schritt für jede ethische Beschäftigung mit Algorithmen. Dies ist gerade in einem Kontext relevant, der sehr stark von Optimierungslogiken geprägt ist. Dennoch muss eine ethische Beschäftigung mit Algorithmen auch über dieses prinzipielle Moment der Offenheit hinaus zu konkreten Alternativen finden können. Es geht in einer Ethik der Algorithmen auch um die Orientierung der Entwicklung anderer oder besserer Verfahren. Auch dabei ist es wichtig, die prinzipielle Unabgeschlossenheit mitzudenken. Und nicht der Verlockung zu verfallen, zu denken, jetzt die richtige Lösung zu entwickeln. Was aber die ethische Fabulation gegenüber etwa derjenigen der Informatik besser macht, bleibt aus Sicht von Amoore Theorie unklar. Die Umsetzung jeglicher Fabulation von Algorithmen wird notwendigerweise auf genau die Kontingenzen stoßen, die Amore ja auch analysiert: Abhängigkeit von Daten, von Code und dessen Möglichkeiten, von materiellen Bedingungen des Algorithmischen (Matzner 2024: 125). Mit diesen muss sich dann im Konkreten und aufgrund von konkreten Kritiken der bestehenden algorithmischen Anwendungen auseinandergesetzt werden. Das bedeutet also nicht, dass Amoore Ethik ihren Zweck verfehlt, aber sie muss durch eine Kritik innerhalb des Algorithmischen ergänzt werden.

3. Der Verlust des Außen

Die dritte Form von Kritik argumentiert in Bezug auf den Verlust des Außen. Auch hier geht es um ein prinzipiell Offenes, Unberechenbares. Wenn aber Amore der Meinung ist, dass gerade auch Algorithmen immer in Bezug zu einem Unabgeschlossenen und Unberechenbaren stehen, so liegt in dieser Form der Kritik die Betonung darauf, dass Algorithmen einer eigenen technischen Logik folgten. Das Offene und Unberechenbare wird auf Seiten der Menschen verordnet und gerade nicht bei den Algorithmen. Die Gefahr besteht dann folglich nicht darin, die prinzipielle Offenheit des Algorithmischen zu vergessen oder zu verleugnen, wie Amore dies schreibt. Stattdessen geht es darum, dass die abgeschlossene Logik des Algorithmischen auf die zwischenmenschliche Offenheit übergreift und diese tatsächlich verdrängt oder zerstört. Diese Möglichkeit von Algorithmen, menschliche Aktion oder Interaktion zu determinieren, wird damit begründet, dass sich die algorithmische Logik dadurch auszeichne, bestimmte menschliche Dispositionen anzusprechen.

Solch ein Argument findet sich z.B. in der Beschreibung der vielzitierten Idee der »Filterblase«. Dieser zufolge kann die algorithmische Auswahl der Inhalte di-

gitaler Medien ein menschliches Grundbedürfnis ansprechen, Dinge, die wir mögen oder die unseren Positionen ähnlich sind, weniger kritisch zu betrachten als andere (Pariser 2011). Im Gegensatz zu der im ersten Abschnitt diskutierten Form der Kritik, die auf autonome Technikkontrolle aufsetzt, wird hier also mit einem Menschenbild operiert, das anerkennt, dass Menschen nicht nur rationale autonome Wesen sind. In der Theorie der Filterblase geht es noch darum, diese Autonomie dennoch so weit wie möglich wieder herzustellen und damit sich gegen den algorithmischen Einfluss zu verwehren. Andere Theorien, die mit dem Verlust des Außen argumentieren, gehen hier weiter und beziehen sich auf ein Menschenbild, das sich von dem liberalen autonomen Subjekt deutlich unterscheidet. In prägnanter Form bringt diese Argumentation zum Beispiel Antoinette Rouvroy in ihrer Konzeption der algorithmischen Gouvernamentalität. Sie erklärt ganz explizit: »I do not intend to rehabilitate the autonomous, unitary, perfectly intentional and rational subject, the fundamental unit of liberalism. As for the ›subject‹ or the ›person‹, I hypothesise that there has never been anything to be nostalgic about.« (Rouvroy 2013: 157) Stattdessen bezieht sie sich auf Louis Althusser, Judith Butler und Jacques Derrida, um das Subjekt zu beschreiben, um das es ihr geht: »These ›pragmatic‹ accounts understand the ›self‹ as a process rather than a phenomenon, a process happening between individuals, in a space that both presupposes and constitutes ›the common‹.« (Rouvroy 2013: 158) Algorithmen bedrohen ihr zufolge also genau diesen Prozess, »the inactual, potential dimensions of human existence, its dimensions of virtuality, the conditional mode of what people ›could‹ do, their potency or agency«. (Rouvroy 2013: 158)

Diese Bedrohung wird möglich, weil Algorithmen die Architekturen und Umgebungen, in denen Menschen handeln, kontrollieren.

»Algorithmic governmentality thus exhibits a new strategy of uncertainty management consisting in minimising the uncertainty associated with human agency: the capacity humans have to do or not to do all they are physically capable of. Effected through the reconfiguration of informational and physical architectures and/or environments within which certain things become impossible or unthinkable, and throwing alerts or stimuli producing reflex responses rather than interpretation and reflection, it affects individuals in their agency [...].« (Rouvroy 2013: 155)

Im Gegensatz etwa zu Foucaultianischen Ansätzen, die Algorithmen als neue Form der Subjektivierung (Matzner 2017) oder Biopolitik beschreiben (Cheney-Lippold 2011), geht es laut Rouvroy gar nicht mehr um einen Einfluss auf Subjekte. Algorithmische Gouvernamentalität muss sich nicht um Subjekte kümmern, weil ihre Absichten, Intentionen und dergleichen gar nicht wichtig sind. Das ist erst einmal eine treffende Beschreibung vieler datengetriebenen Anwendungen. Hier wird da-

mit operiert, kleinste Datenspuren, die aus der Sicht eines Subjektes völlig belanglos sein mögen, mittels Mustererkennung zu verknüpfen. Was zählt, ist nicht eine umfassende digitale Erfassung eines Subjekts, sondern die Sammlung von möglichst vielen »Signalen« unterschiedlicher Subjekte, die das zu operationalisierende Ziel beschreiben. Wenn es also beispielsweise darum geht, potentielle Kund*innen einer Versicherung zu bewerten, geht es nur darum, was in der Breite der Daten z.B. für die Wahrscheinlichkeit eines Schadensfalls spricht. Der Rest des Subjekts ist egal, oder wie Rouvroy schreibt »does not matter«. Die Idee eines digitalen Abbilds des Subjekts, eines »data double« (Lyon 2014), das dann mit dem realen Subjekt kritisch verglichen werden könnte, zielt also an der Realität vorbei.

Wenn Algorithmen kein Subjekt mehr erfassen wollen, versuchen sie auch keines als Subjekt zu beeinflussen. Stattdessen zielen Algorithmen auf Umgebungen und Architekturen, in denen sich Subjekte vermeintlich frei bewegen könnten. Empfehlungssysteme wählen etwa im Hintergrund mediale Inhalte oder Produkte von Onlineshops (inklusive deren Preise) aus, während das für die Kund*innen eben »der Inhalt« des Mediums oder der Suche im Shop ist. Sie werden nicht als Subjekt adressiert, die Inhalte sind einfach da. Dass andere ganz andere Auswahlen bekommen und an welchem Merkmal der eigenen Person das liegt, ist wenn überhaupt nur durch aufwändige Vergleiche nachvollziehbar. Rouvroy hatte beim Schreiben des Textes sicher auch Anwendungen im Blick, die geplant waren, aber immer noch nicht richtig funktionieren. Biometrische Kontrollen öffnen den einen fast magisch Türen, während sie andere diskret aussortieren. Intelligente Videoüberwachung kontrolliert im Hintergrund Passagierflüsse in der U-Bahn und lässt störende Personen entfernen. All diesen Konfigurationen von Umgebungen ist gemein, dass Rouvroy sie mit behavioristischem Vokabular beschreibt. Sie verändern Handlungsmöglichkeiten, nicht durch Ansprache von Subjekten, sondern arbeiten »through the reconfiguration of informational and physical architectures and/or environments within which certain things become impossible or unthinkable, and throwing alerts or stimuli producing reflex responses« – wie oben zitiert. Rouvroy operiert hier mit Denkfiguren, die an die frühe Kybernetik erinnern. Hier wurde das liberale Menschenbild durch die Idee herausgefordert, auch der Mensch sei nur eine Feedbackmaschine. Gleichzeitig war die Kybernetik aber auch ein liberales Projekt, das technischen Fortschritt verfolgte. So schreibt N. Kathrine Hayles über den Gründer der Kybernetik, Norbert Wiener:

»Throughout his major writings, he struggled to reconcile the tradition of liberalism with the new cybernetic paradigm he was in the process of creating. When I think of him, I imagine him laboring mightily to construct the mirror of the cyborg. He stands proudly before this product of his reflection, urging us to look into it so that we can see ourselves as control-communication devices, differing in no substantial regard from our mechanical siblings. Then he happens to glance over

his shoulder, sees himself as a cyborg, and makes a horrified withdrawal.« (Hayles 1999: 87)

Ganz ähnlich wird auch für Rouvroy der Mensch gegenüber dem Algorithmus zu einem Reiz-Reaktionswesen, nur dass hier nicht mehr das liberale Subjekt bedroht ist, sondern die zwischenmenschliche Subjektkonstitution. Was hier verloren geht, beschreibt sie selbst als »Außen«:

»How do we find an ›outside‹, an excess of the world over reality, a space of recalcitrance from which to gain solidity and to practise critique? [...] We] should realise that the fundamental stake – what has to be preserved as a resource antecedent to both the ›subject‹ and sociality, as excess of the world over the algorithmic reality, is ›the common‹; this ›in between‹, this space of common appearance (comparution) within which we are mutually addressed to each other.« (Rouvroy 2013: 160)

Es ist nicht ganz klar, welchen ontologischen Status das menschliche Subjekt in dieser Argumentation hat. Auf der einen Seite könnte diese ganz ähnlich wie die Kritik Amoore gelesen werden. Dann ginge es also darum, dass es im menschlichen Handeln eine prinzipielle Offenheit und konstitutives Aufeinanderbezogen-sein gibt, ein »exzess«, welcher durch algorithmische Anwendungen verdeckt oder ignoriert wird. (Auch wenn der relevante Unterschied zu Amoore bliebe, dass dieses Außen eine rein menschliche Sache ist.) An den Stellen, wo Rouvroy aber von Reiz und Reflex spricht, klingt es so, als sähe sie tatsächlich die Möglichkeit, dass dieses Außen, d.h. menschliche Sozialität durch Algorithmen verschwindet; dass der Mensch eben als Reizreaktionstier anfällig ist für diese Form von Steuerung, in der dann der inhärente Exzess menschlicher Interaktion nicht mehr nur verdeckt oder ignoriert wird, sondern tatsächlich zerstört ist.

Folgt man der zweiten Lesart, so wäre das eine ziemlich technikdeterministische Sicht. Algorithmen könnten dann quasi menschliche Sozialität mit behavioristischer Steuerung überschreiben. Diese Sicht übersieht, dass die Algorithmen nicht so klar und deterministisch funktionieren, wie versprochen. Gerade die unbemerkte, unbewusste Manipulation durch Algorithmen ist ein oft diskutiertes Risiko, das sich aber in der Realität als deutlich komplexer herausstellt, als Menschen nur den richtigen Reiz vorzuwerfen. Hier muss die Kritik also darauf achten, den überzogenen Ansprüchen der Anbieter und Entwickler solcher Software nicht einfach zu folgen. Ein exemplarischer Fall wären die Versprechen von Cambridge Analytica, sogar wichtige Wahlen manipulieren zu können, die sich im Nachhinein als wenig haltbar herausgestellt haben (Denham 2020). Zumindest muss gesehen werden, dass algorithmische Veränderungen von Umgebungen und Handlungsmöglichkeiten nicht einfach Menschen treffen, sondern sozial und kulturell situierte Personen, die entsprechend unterschiedlich auf die algorithmischen Ak-

tivitäten reagieren oder unterschiedlich von ihnen betroffen sind (Matzner 2019a). Diese Differenzen zwischen unterschiedlichen Auswirkungen desselben Algorithmus auf sozio-kulturell verschieden situierte Menschen, bräuchte eben dann eine Kritik von »innen«, das heißt basierend auf der Differenz zwischen verschiedenen sozio-technischen Konfigurationen. Ebenso müssen dann auch die verschiedenen Auswirkungen diverser Algorithmen, die dann nicht mehr alle in dasselbe einfache Reiz-Reaktions-Schema passen, in den Blick genommen werden.

Auch wenn nach der ersten Lesart die Offenheit zwischenmenschlicher Interaktion als prinzipiell gegeben gesehen wird, die dann durch Algorithmen verdeckt oder ignoriert wird, bleibt das Argument in dieser Form bei einer Gegenüberstellung von menschlicher Handlung und technischer Konfiguration von Umgebungen. Für eine ethisch-politische Bewertung fehlt auch hier eine soziale Kontextualisierung der jeweils betroffenen menschlichen Subjekte. Diese läge vor dem theoretischen Hintergrund von Rouvroys Arbeiten sehr nahe. Ihre Quellen wie Althusser, Butler, oder Boltanski haben alle auf unterschiedliche Weise versucht, soziale, ökonomische und kulturelle Differenzen nicht nur in Bezug auf gesellschaftliche Strukturen, sondern auch auf der individuellen Ebene zu fassen. Dass diese Analyse hier fehlt, liegt also nicht am theoretischen Design, sondern eher an einer anderen Interessenlage – vielleicht auch an dem Eindruck, den kybernetische oder behavioristische Beschreibungen vom Verschwinden des Menschen machen können.

4. Schluss

In diesem Text wurden drei unterschiedliche Formen präsentiert, in denen Algorithmen »von außen« kritisiert werden. Die erste Form der Kritik, eine Gegenüberstellung von Mensch und Technik basierend auf einem instrumentellen Technikverständnis, ist schon aus technikphilosophischer Sicht – also noch vor der Frage nach Kritik – vielfältig kritisiert worden. Sie taucht hier dennoch auf, weil sie nach wie vor viele konkrete – und dann auch normative – Vorhaben zum Umgang mit Technik informiert: Gesetze, Data Literacy Workshops, Ethikrichtlinien und mehr. Entsprechend gilt es zu fragen, wie diese Vorhaben zu einem techniktheoretisch angemesseneren Begriff von Technik gebracht werden können.

Für die Frage nach einer Kritik von Algorithmen relevanter sind aber die beiden anderen Formen der Kritik. Beide liefern wichtige Inspirationen, müssen aber durch eine kontextualisierte Kritik »von innen« ergänzt – aber nicht ersetzt – werden. Das »Unberechenbare« als ein »Außen« der Algorithmen zeichnet sich im Gegensatz zu den beiden anderen hier diskutierten Formen dadurch aus, dass es nicht mit einem Gegensatz zwischen Menschen und Algorithmen arbeitet, sondern prinzipielle Unverfügbarkeiten für die Technologie an sich sucht. Es wurde gezeigt, dass die grundlegenden Arbeiten der Informatik, in der Form, wie sie in der aktuellen

Kritik aufgenommen werden, das heißt ohne Einbeziehung der Erkenntnisse zu probabilistischem und statistischem Rechnen, nicht viel zu aktuellen Anwendungen des Maschinellen Lernens und Ähnlichem sagen können. Mit Amoore wurde aber gezeigt, dass die Figur des Unberechenbaren dennoch relevant werden kann. Indem sie Algorithmen als eine Form des Schreibens auffasst, umgeht sie die Einschränkungen, die engere, auf bestimmte Eigenschaften von Informationstechnik ausgerichtete Reflexionen haben. Durch ihre Analyse des Schreibens mit Foucault und Derrida umgeht sie auch die Problematik, die Programmierung als entscheidenden Ort des Schreibens von Algorithmen überzubewerten – oder die Fetischisierung von Code wie Wendy Chun diese Sicht prominent kritisiert hat (Chun 2008). Stattdessen schreiben die Algorithmen selbst und sind damit aber der prinzipiellen Unabgeschlossenheit jeglichen Schreibens ausgesetzt. Kritisch arbeitet diese Unabgeschlossenheit vor allem negativ: gegen eine zu technikdeterministische Sicht, gegen die vermeintliche Objektivität von Algorithmen, gegen die Überbetonung der Macht von Programmierern. Wie diese prinzipielle Unabgeschlossenheit aber jeweils zu füllen ist, wird damit nicht beantwortet. Für Amoores ethnographisch inspirierte Arbeitsweise ist das kein Problem, sie liefert diese Inhalte aus der Empirie. Hier müsste aber nicht nur empirisch, sondern auch normativ angesetzt werden, was dann aber ohne Einbezug der konkreten Anwendung und Situation der Subjekte nicht zu leisten ist.

Rouvroy und ähnlich argumentierende Autor*innen kehren zurück zu einem Gegensatz zwischen Mensch und Technik. Während aber die Seite menschlicher Subjekte hier komplex ko-konstitutiv gedacht ist, wirkt die Technik deterministisch-reduktionistisch. Rouvroy wählt dazu selbst den Begriff des »data behaviorism«. Damit wird implizit auch die sozialtheoretische Komplexität reduziert. Während Menschen in vielerlei Hinsicht aufeinander bezogen sind, und damit sowohl in Machtverhältnissen stehen, als auch die Möglichkeit eines Exzesses, eines Aufbrechens, einer Veränderung besteht, wird diese Differenzierung in beide Richtungen eingeebnet. Gegenüber der Technik wird der Mensch zu einem Reiz-Reaktions-Wesen. Die Bedrohungen, die Rouvroy hier konkret im Kopf hat, sind sicher real, bedürfen aber zweierlei Präzisierung. Inzwischen wurde deutlich, dass Technik nicht alle Menschen gleich bedroht, die vermeintlich basal-behavioristische Wirkweise hat selbst sozio-kulturelle Indizes. Und zweitens funktioniert die Technik nicht so glatt und deterministisch, wie sie hier wirkt. In Bezug auf diesen Punkt wäre es spannend, die zweite Form der Kritik, die eine Unabgeschlossenheit der Technik thematisiert, mit der dritten, welche die Offenheit menschlicher Interaktion betont, zu verbinden. Davon könnte die Amooresche Form der Kritik umgekehrt ein stärker zwischen Technik und Menschen ko-konstitutives und damit auch an sozio-politischer Normativität reichhaltigeres Technikverständnis bekommen.

Literatur

- Amoore, L. (2009): Algorithmic War. Everyday Geographies of the War on Terror, in: *Antipode*, 41(1), 49–69.
- Amoore, L. (2020): Cloud ethics. Algorithms and the attributes of ourselves and others, Durham: Duke University Press.
- Bruns, A. (2019): Are Filter Bubbles Real?, Cambridge: Polity Press.
- Cheney-Lippold, J. (2011): A New Algorithmic Identity. Soft Biopolitics and the Modulation of Control, in: *Theory, Culture & Society*, 28(6), 164–181.
- Chun, W.H.K. (2008): On »sourcery,« or code as fetish, in: *Configurations*, 16(3), 299–324.
- Church, A. (1936): An Unsolvable Problem of Elementary Number Theory, in: *American Journal of Mathematics*, 58(2), 345–363.
- Denham, E. (2020): Letter RE. ICO investigation into use of personal information and political influence, 2020. [https://ico.org.uk/media/action-vevetaken/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf] (Zugriff: 29.05.2024).
- Derrida, J. (2001): Writing and difference, London: Routledge.
- Foucault, M. (1988): Was ist ein Autor?, in: Ders. (Hg.), *Schriften zur Literatur*, Frankfurt a.M.: Suhrkamp, 7–31.
- Hayles, N.K. (1999): How We Became Posthuman, Chicago: University of Chicago Press.
- Kitchin, R. (2017): Thinking Critically about and Researching Algorithms, in: *Information, Communication & Society*, 20(1), 14–29.
- Krämer, S. (2021): Digitalism as a Cultural Technique. From Alphanumeric to AI, in: *Goethe.De.*, 2021. [<https://www.goethe.de/prj/k40/en/eth/dig.html>] (Zugriff: 26.05.2024).
- Lyon, D. (2014): Surveillance, Snowden, and Big Data. Capacities, Consequences, Critique, in: *Big Data & Society*, 1(2), 1–13.
- Matzner, T. (2017): Opening Black Boxes Is Not Enough. Data-Based Surveillance In Discipline and Punish And Today, in: *Foucault Studies*, 23, 27–45.
- Matzner, T. (2019a): Plural, situated subjects in the critique of artificial intelligence, in: Sudmann, A. (Hg.), *The Democratization of Artificial Intelligence*, Bielefeld: transcript, 109–121.
- Matzner, T. (2019b): The Human Is Dead – Long Live the Algorithm! Human-Algorithmic Ensembles and Liberal Subjectivity, in: *Theory, Culture & Society*, 36(2), 123–144.
- Matzner, T. (2022): Algorithms as Complementary Abstractions, in: *New Media & Society*, 26(4). [<https://doi.org/10.1177/14614448221078604>].
- Matzner, T. (2024): Algorithms. Technology, Culture, Politics, Abingdon/Oxon: Routledge.

- Matzner, T.; Masur, P.K.; Ochs, C.; von Pape, T. (2016): Do-It-Yourself Data Protection – Empowerment or Burden?, in: Gutwirth, S.; Leenes, R.; De Hert, P. (Hg.), *Data Protection on the Move*, Dordrecht: Springer, 277–305.
- Mühlhoff, R. (2019): Menschengestützte Künstliche Intelligenz. Über die soziotechnischen Voraussetzungen von »deep learning«, in: *Zeitschrift für Medienwissenschaft*, 11(2), 56–64.
- Pariser, E. (2011): *The Filter Bubble. What the Internet Is Hiding from You*, London: Viking.
- Parisi, L. (2016): Automated Thinking and the Limits of Reason, in: *Cultural Studies ↔ Critical Methodologies*, 16(5), 471–481.
- Rohlfing, K.J.; Cimiano, P.; Scharlau, I.; Matzner, T.; Buhl, H.M.; Buschmeier, H.; Esposito, E. et al. (2020): Explanation as a social practice. Toward a conceptual framework for the social design of AI systems, in: *IEEE Transactions on Cognitive and Developmental Systems*, 13(3), 717–728.
- Rouvroy, A. (2013): The end(s) of critique. data-behaviourism vs. due-process, in: De Vries, K.; Hildebrandt, M. (Hg.), *Privacy, Due Process and the Computational Turn. The Philosophy of Law Meets the Philosophy of Technology*, London: Routledge, 143–167.
- Seyfert, R.; Roberge, J. (Hg.) (2016): *Algorithmic Cultures. Essays on Meaning, Performance and New Technologies*, London/New York: Routledge.
- Turing, A.M. (1937): On Computable Numbers, with an Application to the Entscheidungsproblem, in: *Proceedings of the London Mathematical Society* 2, 42(1), 230–265.
- Zuboff, S. (2016): Wie wir Googles Sklaven wurden, in: *FAZ.NET*, 05.03.2016. [<https://www.faz.net/aktuell/feuilleton/debatten/die-digital-debatte/shoshana-zuboff-googles-ueberwachungskapitalismus-14101816.html>] (Zugriff: 26.05.2024).
- Zuboff, S. (2019): *The Age of Surveillance Capitalism. The Fight for the Future at the New Frontier of Power*, London: Profile Books.

