Antonia Kempkens

Zwischen Nutzen und Risiko

Den Umgang mit digitalen Daten ethisch gestalten

VERLAG KARL ALBER

Ethics, Law and AI

Herausgegeben von

Carmine Di Martino (Università degli Studi di Milano)
Federico L.G. Faroldi (Università di Pavia)
Roberto Redaelli (Università degli Studi di Milano)

Band 5

Antonia Kempkens

Zwischen Nutzen und Risiko

Den Umgang mit digitalen Daten ethisch gestalten

VERLAG KARL ALBER

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

1. Auflage 2026

Zugl.: Bremen, Univ., Diss., 2026
u.d.T.: Zwischen Nutzen und Risiko:
den Umgang mit digitalen Daten ethisch gestalten

Publiziert von
Verlag Karl Alber – ein Verlag in der
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.**verlag-alber.de**

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-495-98772-8
ISBN (ePDF): 978-3-495-98773-5

DOI: https://doi.org/10.5771/9783495987735

Onlineversion
Inlibra

Danke an Prof. Dr. Dagmar Borchers und Prof. Dr. Dr. Björn Niehaves für die Betreuung

Danke an das Unternehmen Dataport für die finanzielle Unterstützung

Danke an Dr. Derek Meier und Tina Siegfried für den Dialog

Danke an Prof. Dr. Georg Mohr für das Feedback

Danke an meine Kolleg:innen für die Gespräche

Danke an das WOC Graduiertennetzwerk für den Austausch

Danke an meine Freund:innen für ihr offenes Ohr

Danke an meine Eltern und Paulina für alles

Danke an Dominik für den Beistand

Danke an Oma

Inhaltsverzeichnis

1 Aktueller Stand der Digitalethik

Alle kennen es: Man möchte einen neuen Ausweis beantragen oder den Wohnsitz ummelden und muss lange auf einen Termin warten. Dann muss man zum Bürgeramt und dort wieder warten. Das soll durch die Digitalisierung der Verwaltungsprozesse bald schneller gehen. Durch das Gesetz zur Verbesserung des Onlinezugangs zu Verwaltungsleistungen (Onlinezugangsgesetz – OZG) sind „Bund und Länder (...) verpflichtet, ihre Verwaltungsleistungen auch elektronisch über Verwaltungsportale anzubieten."[1] Das heißt, dass es möglich sein soll, Verwaltungsleistungen online zu beantragen, ohne auf einen Termin beim Amt warten zu müssen. Dies wird durch die Digitalisierung ermöglicht. Im Englischen werden zwei Dimensionen der Digitalisierung unterschieden: Während „Digitization" die technischen Möglichkeiten der Digitalisierung beschreibt und den Einsatz von digitalen Technologien zur Umwandlung von analogen Werten in digitale Formate bezeichnet[2], beschreibt „Digitalization" den Prozess der Digitalisierung, also die Einführung und verstärkte Nutzung dieser technischen Möglichkeiten.[3] Diese verstärkte Nutzung der technischen Möglichkeiten („Digitalization") soll auch – auf der Grundlage des Onlinezugangsgesetzes (OZG) in die Verwaltung eingeführt und dort verstärkt genutzt werden. Dies fußt auf der „Digitization", welche es ermöglicht, Daten digital zu speichern und schneller zu verarbeiten. So soll zum Beispiel das Beantragen von Wohngeld durch den Einsatz digitaler Technologien von Zuhause aus möglich sein.

Um solche digitalen Anträge möglich zu machen, müssen sogenannte Datenökosysteme gestaltet werden. Datenökosysteme können als komplexe sozio-technische Systeme definiert werden, welche

1 §1a, Abs. 1, Satz 1 OZG
2 Donges 2022:212f.; Görs et al. 2022:12; Klenk et al. 2020:5; Altintac 2022:3,6; Corsten, Roth 2022:10; Harwardt 2022:3; Traum et al. 2017:2
3 Görs et al. 2022:12; Traum et al. 2017:2f.

durch das Zusammenspiel zwischen menschlichen Akteuren[4] und Technologien das Sammeln, Speichern und Verarbeiten von Daten sowie das Bereitstellen der Ergebnisse ermöglichen.[5] Im Folgenden wird ausführlich definiert, was ein Datenökosystem ist, und es wird ein Modell von der Struktur eines Datenökosystems erstellt, in welchem der Datenfluss zwischen vier Akteuren abgebildet wird: Datenquellen, Datengenerierenden, Datenverarbeitenden und Datennutzenden. Teil eines Datenökosystems zu sein kann viele Vorteile haben, aber auch Risiken bergen. In dieser Arbeit soll der Umgang mit digitalen Daten in Datenökosystemen systematisch beleuchtet werden.

Die Vorteile und Risiken von Datenökosystemen lassen sich anhand von Beispielen aus der öffentlichen Verwaltung einführend aufzeigen: Bei der digitalen Gestaltung des Wohngeldantrages in Form eines Datenökosystems, geben Antragstellende ihre Daten online an. Um einen Antrag auf Wohngeld stellen zu können, müssen viele persönliche Informationen preisgegeben werden, wie zum Beispiel Name, Anschrift, Einkommensnachweise, gegebenenfalls der Nachweis einer Schwerbehinderung sowie Bescheide über Kindergeld, Kinderzuschlag oder Elterngeld. Die Wohngeldbehörde speichert und verarbeitet diese Daten, um feststellen zu können, ob die Antragstellenden Anspruch haben auf Wohngeld und der Staat verpflichtet ist, ihnen dies auszuzahlen. Nach der Prüfung bekommen die Antragstellenden Bescheid, ob sie Anspruch auf Wohngeld haben oder nicht. Gleichzeitig werden die Daten anonymisiert (ohne Namen und Anschrift) an das Statistische Landesamt, an das Statistische Bundesamt sowie an das Bundesministerium des Innern, für Bau und Heimat und an das Bundesamt für Bauwesen und Raumordnung weitergegeben, damit diese die Daten im Rahmen einer Wohngeldstatistik verarbeiten können.[6]

Der Vorteil der digitalen Gestaltung des Wohngeldantrages in einem Datenökosystem liegt auf der Hand: ein geringerer Arbeitsaufwand für die Antragstellenden und die Wohngeldbehörde. Für die Antragstellenden ist es weniger Aufwand, ihre Daten online

4 Mit dem Begriff „Akteur" sind hier und im Folgenden Personen aller Geschlechter gemeint.

5 Oliveira et al. 2019:590

6 §§ 34–36 WoGG; Siehe zum Beispiel der Wohngeldantrag des Landes Bremen: https://www.wohngeld.org/wp-content/uploads/AntragMZ_LZ_2021.pdf

anzugeben als auf einen Termin in der Behörde zu warten und dort hinzufahren. Die Mitarbeitenden der Wohngeldbehörde haben weniger Aufwand, weil sie die Dateneingabe einsparen, alle notwendigen Dokumente digital abrufen können und bei der Auswertung durch einen Algorithmus unterstützt werden. Ein solches digitales Datenökosystem kann allerdings auch Risiken bergen. Zum einen werden diese personenbezogenen Daten digital gespeichert und dadurch dem Risiko der Zweckentfremdung ausgesetzt. Zum anderen erzeugt der Einsatz von Algorithmen bei der Anspruchsbewertung das Risiko, dass die Entstehung der Ergebnisse für Antragstellende schwer oder gar nicht nachvollziehbar ist. Es besteht die Gefahr, dass Antragstellende nicht nachvollziehen können, warum ihr Antrag abgelehnt wurde und nicht mehr überblicken, für welche Zwecke ihre Daten ansonsten verarbeitet werden.

Diese Risiken lassen sich auch in Projekten des Bundes identifizieren. So soll zum Beispiel das vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) geförderte Projekt MERLOT (MarkEtplace foR LifelOng educaTional dataspaces and smart service provisioning) Bildungsdatenräume und -dienste in Form von einem Datenökosystem erschaffen.[7] Dabei geht es unter anderem darum, Lernende bei der Bildungs- und Berufsorientierung zu unterstützen. Hierfür sollten ab 2022 digitale Karriereorientierungsassistenten, ein Weiterbildungsassistent und ein Bildungsprofilprognoseassistent entwickelt werden. Zu dem Zweck sollten Daten über reale Bildungswege erhoben werden und schon vorhandene Bildungsprofildaten aus unterschiedlichen Quellen zugänglich gemacht werden. Diese Bildungsdaten sollten mit Hilfe von Algorithmen verarbeitet werden, mit dem Anspruch, den Datenschutz, die Datensouveränität[8] der Datenquellen nicht zu gefährden. Mit Hilfe des Einsatzes von Künstlicher Intelligenz (KI) sollten die kognitiven Positionen der Lernenden dem (Weiter-) Bildungsangebot zugeordnet werden. Das Ergebnis sollten Hinweise für Lernende sein, wel-

7 Mir liegt die Gesamtvorhabenbeschreibung aus dem Jahr 2021 vor auf Grund von internem Zugang durch Dataport (Lehmann et al. 2021). Der Abschlussbericht liegt mir jedoch nicht vor, weshalb die Schritte des Projektes als Vorhaben beschrieben werden.

8 Datensouveränität bedeutet in dem Fall, dass diejenigen, über die die Daten eine Aussage machen, der Verwendung ihrer Daten zustimmen oder diese ablehnen können und somit die „Souveränität" über ihre Daten haben.

che Bildungsangebote oder Berufswege ihnen zur Verfügung stehen. Anschließend sollten neue KI-Systeme mit Hilfe der erhobenen Bildungsdaten trainiert werden.[9]

Der Schutz der Bildungsdaten wird in der Vorhabenbeschreibung garantiert, aber gleichzeitig werden Bildungsdatenräume geschaffen, damit die Bildungsdaten zu verschiedenen Zwecken verwendet werden können – zum Beispiel zum Trainieren von KI-Systemen oder zum Zuordnen der kognitiven Positionen der Lernenden zu dem (Weiter-)Bildungsangebot. Damit die Lernenden, deren Daten erhoben wurden, ihre Datensouveränität erhalten können, sollte ihre Zustimmung zu der Verarbeitung ihrer Daten zweckbezogen eingeholt werden und widerrufbar sein. Auf diese Weise wird das Risiko der Zweckentfremdung aktiv thematisiert und minimiert. Was jedoch nicht thematisiert wird, ist das Risiko der fehlenden Nachvollziehbarkeit der Ergebnisse. Denn wenn eine Person einen der digitalen Assistenten nutzt und eine Empfehlung für eine Weiterbildung oder einen Berufsweg bekommt, sollte es für sie nachvollziehbar sein, warum sie *diese* Empfehlung bekommt und nicht eine andere. Da die Empfehlungen durch die Verarbeitung von Bildungsdaten anderer Personen mit Hilfe von KI-Systemen generiert werden, muss man sich fragen, welche Aussagen durch die Analyse von Bildungsdaten anderer überhaupt gemacht werden können. Angedeutet werden soll an dieser Stelle schon mal, dass dieses Vorgehen das Risiko birgt, Bestehendes zu replizieren, ohne dass dies durchsichtig wird. Diese Frage wird im Folgenden ausführlich erörtert (Kapitel 7). Hier soll aber schon deutlich werden, dass die Gestaltung von Datenökosystemen in der öffentlichen Verwaltung nicht nur Vorteile mit sich bringt, sondern auch Risiken birgt – ethische Risiken.

Wie genau ethische Risiken zu definieren sind, wird im Verlauf dieser Arbeit geklärt. An dieser Stelle soll nun aber begründet werden, warum das Digitalisieren der Datenökosysteme in der Verwaltung so einer kritischen Betrachtung unterzogen werden muss. Denn man könnte argumentieren, dass beim Beantragen von Verwaltungsleistungen auch jetzt schon analoge Datenökosysteme bestehen. Die Antragstellenden müssen vor Ort beim Amt die gleichen sensiblen Daten angeben, wie in einem Online-Antrag. Die Wohngeldbehörde muss die Daten auch bearbeiten und dann abheften, um bei

9 Lehmann et al. 2021

Beschwerden oder Fehlern eventuell darauf zurückgreifen zu können. Zum Schluss wird den Antragstellenden ebenfalls mitgeteilt, ob sie ein Anrecht auf Wohngeld haben oder nicht. Der Unterschied bei einem digitalen Datenökosystem ist aber, dass die Daten nicht analog abgeheftet werden und dementsprechend in einem ganz bestimmten Aktenschrank in einer konkreten Wohngeldbehörde zu finden sind, sondern sie werden digital gespeichert und sind dementsprechend prinzipiell von überall aus zugänglich. Natürlich sollen die Daten nicht von überall für jeden zugänglich sein, weshalb sie geschützt werden – leider reicht aber der Schutz der Daten oft nicht aus.

Auf Grund menschlicher Fehler und der Komplexität der Software gibt es oft inhärente Schwachstellen in einem System oder seinem Design, die es Unbefugten ermöglichen, Anweisungen auszuführen, sich ohne entsprechende Berechtigung Zugang zu Daten zu verschaffen und Angriffe durchzuführen. Angriffe bezeichnen vorsätzliche Handlungen, die darauf abzielen, einem System Schaden zuzufügen oder seinen regulären Betrieb zu unterbrechen, indem Schwachstellen durch verschiedene Taktiken und Werkzeuge ausgenutzt werden. So kann zum Beispiel Schadsoftware in ein System gebracht werden, indem interne Mitarbeitende zum Betrachten von Phishing Mails und dem Öffnen deren Anhänge bewegt werden. Dadurch können sich die Angreifenden Zugang zum System und den darin enthaltenen Daten verschaffen. Alternativ kann durch massenhaftes Aufrufen eines Systems der Zugang zum System blockiert werden (Denial-of-service) oder das System wird anderweitig gestört bzw. zerstört. Angreifende können einzelne Personen mit Fachkenntnissen im Bereich Hacking sein oder aber Banden des organisierten Verbrechens, und auch staatliche Akteure können hinter Angriffen stecken. Die Angreifenden können unterschiedliche Motive und Ziele haben, wie zum Beispiel finanziellen Gewinn, Beeinflussung der öffentlichen Meinung, oder Spionage. Sobald sie aber Zugang haben zu personenbezogenen Daten, werden auch Identitätsdiebstahl oder Erpressung möglich.[10]

Leider gibt es dafür viele aktuelle Beispiele mit fatalen Auswirkungen: Im Jahr 2017 wurde Equifax, eines der weltweit größten Finanzdienstleistungsunternehmen, gehackt. Dabei wurden die persönli-

10 Abomhara, Køien 2015:71–74; Sasi et al. 2024:457

chen Daten von fast 143 Millionen Kund:innen erbeutet, darunter Sozialversicherungsnummern, Wohnadressen und Geburtsdaten.[11] Die Britische Post (Royal Mail) hatte Anfang 2023 für 6 Wochen Probleme mit dem Versand von Briefen und Paketen nach Übersee auf Grund eines Cyberangriffs auf die Computersysteme, welche die Britische Post für den Versand von Post ins Ausland nutzt.[12]

Vastaamo war Finnlands größtes Psychotherapieunternehmen mit jährlich weit über 10.000 Kunden. Es wurde zunächst als Telegesundheitsdienst entwickelt und bot später auch psychotherapeutische Behandlung über eine Plattform an. Im Oktober 2020 wurde bekannt, dass Vastaamo von einer Datenpanne betroffen war, bei der es dem Angreifer gelang, persönliche Informationen über mehr als 33.000 seiner Patienten und deren psychotherapeutische Behandlungen zu stehlen – darunter auch vertrauliche Therapieunterlagen, wie zum Beispiel Notizen aus den Therapiesitzungen. Der Hacker kontaktierte den CEO von Vastaamo erstmals am 28. September 2020 mit einer Lösegeldforderung in Höhe von 40 Bitcoin (damals entsprach das einem Wert von rund 450.000 Euro). Er behauptete, dass er jeden Tag 100 Datensätze durchsickern lassen würde, bis das Lösegeld eingegangen sei. Da Vastaamo den Lösegeldforderungen nicht nachkam, verlagerte der Angreifer seine Bemühungen auf die Erpressung der Patient:innen, deren Daten gestohlen wurden. Es wurden ca. 30.000 der Patient:innen angeschrieben und ein Lösegeld von 200 bis 500 Euro gefordert. Um seine Drohung zu verstärken, veröffentlichte der Angreifer Patientenakten und Notizen von Sitzungen von mehr als 2.000 Vastaamo-Kunden im Darknet.[13] Laut der finnischen Tageszeitung Helsingin Sanomat hat ein Rechtsvertreter, der etwa 1.500 Opfer vertritt, erklärt, dass sich mehrere der Vastaamo Kund:innen, deren Therapienotizen veröffentlicht wurden, daraufhin das Leben genommen haben.[14] Hier wird deutlich, wie fatal die Zweckentfremdung von personenbezogenen Daten sein kann. Zusätzlich zu dem Schmerz, der durch die Weitergabe hochsensibler Informationen verursacht wurde, sind die Vastaamo Kund:innen auch anfälliger für mögliche Identitätsdiebstähle in der

11 Durnell et al. 2020:1836
12 Espiner, Tidy 2023; Sweney 2023
13 Looi et al. 2025:107f.; Ghanbari, Koskinen 2024:1f., 4ff.
14 Helsingin Sanomat 2024; Ghanbari, Koskinen 2024:2

Zukunft, da ihre persönlichen Informationen im Darknet verfügbar waren.[15]

Solche und ähnliche Datenpannen und Cyberattacken werden im DSGVO-Portal gesammelt[16] – es sind oft auch Einrichtungen der öffentlichen Verwaltung betroffen. So wurde zum Beispiel am 25. April 2025 die digitale Infrastruktur der Stadt Berlin lahmgelegt. Es gab einen Denial-of-service Angriff auf die IT-Systeme der Berliner Verwaltung, indem eine Flut künstlich generierter Anfragen die Onlineportale der Stadt für Nutzende unerreichbar machten. In diesem Fall wurden keine Daten gestohlen, doch auch das kommt regelmäßig vor.[17] Mitte April 2025 hat die Safepay-Ransomware-Gruppe behauptet, 85 GB Daten aus den Verwaltungssystemen der Stadt Heilbronn erbeutet zu haben und drohte diese zu veröffentlichen.[18] Am 28. Januar 2025 hat die Ransomware-Gruppierung „Cloak" tatsächlich einen Datensatz veröffentlicht, den sie nach eigener Angabe von der Webseite der Gemeinde Kaisersbach in Baden-Württemberg gestohlen hatten.[19] Das bedeutet, wenn mehr Daten im Verwaltungskontext digital gespeichert werden, sind bei Städten und Gemeinden noch mehr sensible Daten zu holen, und die Angriffe werden noch attraktiver für die Angreifenden.

Bei der digitalen Gestaltung von Verwaltungsleistungen in Form eines Datenökosystems, werden sensible Daten digital gespeichert und sind dadurch dem Risiko einer Zweckentfremdung ausgeliefert. Hinzu kommt, dass diese Daten mit Hilfe von Algorithmen verarbeitet werden. Dadurch können im Verwaltungskontext mehr Anträge schneller bearbeitet werden, sodass die Antragstellenden früher ihren Bescheid bekommen – ein klarer Vorteil. Im Fall eines Wohngeldantrages sind die Bedingungen für eine Bewilligung des Geldes gesetzlich eindeutig vorgegeben – hier spielen Faktoren wie die Höhe des Einkommens und die Größe der Wohnung eine Rolle. Deshalb würden in diesem Kontext nur Wenn-Dann-Algorithmen

15 Ghanbari, Koskinen 2024:6
16 https://www.dsgvo-portal.de/sicherheitsvorfall-datenbank/
17 https://www.dsgvo-portal.de/sicherheitsvorfaelle/ddos_angriff_stadt-berlin-9278.php
18 https://www.dsgvo-portal.de/sicherheitsvorfaelle/ransomware_angriff_stadt-heilbronn-9177.php
19 https://www.dsgvo-portal.de/sicherheitsvorfaelle/ransomware_angriff_gemeinde-kaisersbach-8211.php

eingesetzt, welche diese klaren Vorgaben in den Anträgen überprüfen. Es gibt aber auch viele andere Beispiele, bei denen die Bedingungen nicht so deutlich festgelegt oder definierbar sind. Wenn, wie bei MERLOT, (Weiter-) Bildungsangebote aus Bildungswegdaten generiert werden sollen, gibt es neben Faktoren wie Ausbildung und Berufserfahrung einen gewissen Ermessensspielraum. Um trotzdem Empfehlungen aussprechen zu können, werden selbstlernende KI-Algorithmen eingesetzt. Dies kann jedoch zu systematischen Fehlern führen, die zunächst schwer zu identifizieren sind.

So hat zum Beispiel das Technologieunternehmen Amazon 2014 für den Zeitraum von einem Jahr bei seinem Einstellungsverfahren ein KI-System eingesetzt. Das KI-System wurde anhand von zuvor eingereichten und bewerteten Lebensläufen trainiert. Auf diese Weise lernte der Algorithmus, welche Bewerbungen in der Vergangenheit erfolgreich gewesen waren, und er identifizierte bestimmte Merkmale dieses Erfolgs, diese wurden dann zu seinen Entscheidungskriterien. In der Vergangenheit waren es vor allem Männer, die im Bewerbungsverfahren erfolgreich waren. Dies hatte zur Folge, dass Bewerbungen von Frauen vom Algorithmus abgelehnt wurden – sogar wenn das Geschlecht nicht direkt genannt wurde, sondern nur durch zum Beispiel eher Frauen zugeordnete Hobbys erkennbar war.[20]

In Österreich wurden im Jahr 2020 die Chancen auf dem Arbeitsmarkt von Jobsuchenden mit Hilfe eines Algorithmus bewertet. Danach wurden die Jobsuchenden in Gruppen eingeteilt, um Fördermaßnahmen gezielt an diejenigen mit mittelguten Chancen zu vergeben. Dabei wurden Frauen, unabhängig von ihrer Ausbildung, geringere Chancen am Arbeitsmarkt zugesprochen als Männern. Zudem wurden Frauen schlechtere Chancen ausgerechnet, wieder Arbeit zu finden, wenn sie Kinder zu betreuen hatten, während die Anzahl von Kindern auf die Chancen von Männern am Arbeitsmarkt keine Auswirkungen hatte. Die Folge dieser ungleichen Einteilung war, dass Frauen auf Grund ihres Geschlechts von Fördermaßnahmen ausgeschlossen wurden. Der Einsatz dieses Algorithmus ist noch im Jahr 2020 gerichtlich verboten worden und wird weiterhin geprüft.[21]

20 Stahl et al. 2023:10f.

21 Wimmer 2019; Köver 2019; Pflügl 2024

In Bezug auf die Digitalisierung von Daten und deren algorithmische Verarbeitung gibt es in der wissenschaftlichen Literatur ausführliche ethische Debatten. Die digitale Ethik ist der Zweig der Ethik, der sich mit der Frage befasst, was wir tun sollen angesichts der Konfrontation mit Digitalisierung und ihren Entwicklungen.[22] Der praktische Philosoph und Wirtschaftsinformatiker Rafael Capurro formuliert das so: „Digital ethics undertakes a critical reflection about ourselves in a world shaped by digital technology."[23] Diese Reflexion über die Menschen im Kontext von technologischen Entwicklungen gibt es laut Capurro schon seit den 1940er Jahren und nannte sich zuerst Computerethik.[24] Dieses Feld wurde zwischendurch auch als Informationsethik bezeichnet und nennt sich heute meist Digitalethik. Der Begriff der Digitalethik soll die Aufmerksamkeit weg von den technologischen Entwicklungen selber und hin zu ihrer Nutzung lenken. Hierunter fallen auch konkretere Bereiche wie zum Beispiel Datenethik oder KI-Ethik.[25]

Die Technologiephilosoph:innen und Digitalethiker:innen Luciano Floridi und Mariarosaria Taddeo unterteilen die ethischen Herausforderungen der Digitalisierung in drei Achsen: Datenethik, Algorithmus-Ethik und Ethik der Praktiken. Demnach liegt der Schwerpunkt der Datenethik auf ethischen Problemen bei der Sammlung und Analyse großer Datenmengen – hierdurch werden philosophische Themen wie Privatsphäre, Vertrauen und Transparenz relevant. Bei der Algorithmus-Ethik geht es vorrangig um die Frage, wer für Fehler, wie zum Beispiel Diskriminierung durch algorithmische Ergebnisse, zur Rechenschaft gezogen werden kann, sodass hier der philosophische Begriff der Verantwortung eine große Rolle spielt. Die Ethik der Praktiken beschäftigt sich damit, wie Datenprozesse gestaltet werden sollten.[26] In dieser Arbeit sollen diese drei Bereiche eine Rolle spielen, um Datenökosysteme umfassend ethisch bewerten zu können.

In der wissenschaftlichen Literatur werden verschiedene Zugänge gesucht zu einer systematischen Digitalethik. Der Medientheoretiker

22 Öhman, Watson 2019:2
23 Capurro 2017:277
24 Ebd.
25 Müller 2023:3
26 Floridi, Taddeo 2016:3

Lev Manovich identifiziert drei Gruppen von Menschen im Kontext von Datenverarbeitung: diejenigen, die Daten erzeugen (sowohl bewusst als auch durch das Hinterlassen digitaler Fußabdrücke), diejenigen, die über die Mittel verfügen, sie zu sammeln, und diejenigen, die über das Fachwissen verfügen, sie zu analysieren.[27] Die Gruppe derjenigen, die die Daten analysieren, ist die kleinste und die privilegierteste, da sie die Regeln für die Nutzung der Daten und die Teilnahme daran bestimmen können.[28] Den Digitalethikern Carl Öhman und David Watson zufolge hängen die ethischen Auswirkungen der Digitalisierung von den Zielen ihrer Entwickler:innen und Nutzenden ab. Digitalethik ermöglicht die Betrachtung von widersprüchlichen Werten und das Ermitteln von Kompromissen. Dadurch kann die digitale Ethik theoretische Analysen von grundlegenden ethischen Fragen – nach zum Beispiel Autonomie, Privatsphäre und Gerechtigkeit – in praktikable Leitlinien für die Gestaltung und Nutzung digitaler Technologien umsetzen.[29]

Es gibt bereits viele solcher ethischen Leitlinien, sowohl von staatlichen als auch privaten Akteuren. Die Digitalethiker:innen Anna Jobin und Marcello Ienca sowie die Bioethikerin Effy Vayena haben im Jahr 2019 84 bestehende KI-Leitlinien analysiert und große inhaltliche Übereinstimmung, jedoch Unterschiede bei den Lösungsansätzen festgestellt. In mehr als der Hälfte der analysierten Leitlinien wird Bezug genommen auf die ethischen Prinzipien Transparenz, Gerechtigkeit, Unparteilichkeit, Verantwortung und Privatsphäre – hier herrscht Einigkeit, dass diese Werte beim Einsatz von KI relevant sein sollten.[30] Es geht in den Leitlinien hauptsächlich darum, Schaden zu verhindern und Werte, wie Privatsphäre, Menschenwürde und Autonomie, zu wahren, anstatt sie aktiv zu fördern. In dieser Arbeit soll es deshalb neben der Bewahrung auch bewusst um die Förderung der relevanten Prinzipien gehen.

27 Manovich 2012:470

28 Boyd, Crawford 2012:675

29 Öhman, Watson 2019:1, 3

30 Werte sind orientierungsbietende Ideale, die als moralisch gut und dementsprechend als wünschens- und erstrebenswert wahrgenommen werden (Friedman et al. 2008:70f.). In Kapitel 6 und 7 dieser Arbeit werden Privatsphäre und Transparenz als solche Werte im Kontext von Datenökosystemen etabliert, indem begründet wird, warum sie wünschens- und erstrebenswert sowie moralisch gut sind.

Ein großes Problem bei der Umsetzung von Leitlinien der digitalen Ethik sind Prinzipien, die miteinander im Konflikt stehen. Jobin et al. haben in ihrer Studie festgestellt, dass in der Praxis Unsicherheit bezüglich der Wahl der Prinzipien besteht, insbesondere wenn es widersprüchliche Prinzipien gibt.[31] Prinzipien können im Konflikt stehen, wenn eines zugunsten der Erfüllung des anderen aufgegeben werden muss. Wenn zum Beispiel eine Wohngeldstatistik erstellt werden soll, um an den bedürftigsten Orten am schnellsten bezahlbaren Wohnraum schaffen zu können, dann braucht man die Zustimmung der Wohngeldempfänger:innen, ihre Daten zu diesem Zweck zu verwenden. Wird die Zustimmung vorausgesetzt und zur Bedingung für die Antragstellung gemacht, so wird die Autonomie der Antragstellenden eingeschränkt. Ist die Zustimmung jedoch freiwillig, kann mit hoher Wahrscheinlichkeit keine repräsentative Wohngeldstatistik erstellt werden und somit können die Baugenehmigungen für bezahlbare Wohnungen in Deutschland nicht gerecht verteilt werden. Hier steht die Autonomie der Wohngeldempfänger:innen im Konflikt mit dem Grundsatz der Verteilungsgerechtigkeit.

Den Technologie- und Wissenschaftsphilosoph:innen Jess Whittlestone, Rune Nyrup, Anna Alexandrova und Stephen Cave zufolge können solche Wertekonflikte nur dadurch gelöst werden, dass einem der Werte Vorrang gewährt wird. Diese Entscheidung lässt sich nur dann begründen, wenn die konfligierenden Werte klar definiert sind, wenn klar ist, was technisch möglich ist, um die Werte zu schützen und wenn die Bedürfnisse der betroffenen Gemeinschaften bekannt sind.[32]

In der Realität aber bleibt es oft der Interpretation überlassen, was die Prinzipien besagen, insbesondere weil ein Prinzip unterschiedliche Bedeutungen haben kann. So enthielten zwar 73 der von Jobin et al. analysierten Leitlinien den Grundsatz „Transparenz", aber dieser hatte verschiedene Bedeutungen von Erklärbarkeit über Explizierbarkeit, Verständlichkeit, Interpretierbarkeit, Kommunikation bis zu Offenlegung.[33] Die Lösungsansätze für die Wahrung von Transparenz beziehen sich in vielen der analysierten Leitlinien auf

31 Jobin et al. 2019:396
32 Whittlestone et al. 2019:198f.
33 Jobin et al. 2019:391, 395

eine verstärkte Offenlegung von Informationen durch diejenigen, die KI-Systeme entwickeln oder einsetzen. Aber die Anforderungen, welche Informationen offengelegt werden sollten, unterscheiden sich stark. Manche fordern Informationen über die Verwendung der KI, andere wollen Zugriff auf den Quellcode, um Transparenz zu ermöglichen. Auch die Offenlegung der Verwendung von Daten wird gefordert oder eine Evidenzbasis für die Verwendung von KI, sowie eine Antwort auf die Frage, wer Verantwortung trägt für die KI.[34] In dieser Arbeit werden die relevanten Prinzipien und Forderungen nach ihrer Umsetzung deshalb ausführlich definiert, um Missverständnissen vorzubeugen.

An dieser Stelle fragt sich, woher man weiß, welche Prinzipien relevant sind. Ethisch relevante Prinzipien werden normalerweise aus ethischen Theorien abgeleitet – die drei meist genutzten ethischen Theorien sind der Konsequentialismus, die Deontologie und die Tugendethik. Ein Beispiel von hieraus abgeleiteten Prinzipien sind die Prinzipien der biomedizinischen Ethik der Philosophen Tom Beauchamp und James Childress: Achtung der Autonomie, Wohltätigkeit, Nicht-Schädigung und Gerechtigkeit.[35] Diese Prinzipien wurden sowohl aus einem konsequentialistischen als auch aus einem deontologischen Ansatz abgeleitet und sollten ursprünglich im bioethischen Diskurs und in der bioethischen Praxis eine Grundlage für ethische Urteile bilden.[36] Heutzutage werden sie auch im Kontext der digitalen Ethik verwendet.[37] Dem Datenethiker Brent Mittelstadt zufolge lassen sich die Grundsätze der biomedizinischen Ethik jedoch nicht einfach auf die digitale Ethik übertragen.[38] Diese Debatte ist zwar interessant, doch in dieser Arbeit werden die im Datenökosystem relevanten Prinzipien aus einer der ethischen Theorien selber abgeleitet, sodass die Prinzipien der biomedizinischen Ethik weiter keine Rolle spielen.

Die KI-Ethikerinnen Larisa Bolte und Aimee van Wynsberghe teilen die Debatten zur KI-Ethik in drei Wellen ein: Während sich die erste Welle auf abstrakte und nicht-existierende Technologie konzen-

34 Jobin et al. 2019:391
35 Beauchamp, Childress 1979
36 Beauchamp, Childress 2019:9
37 Floridi, Cowls 2021; Jobin et al. 2019; Ryan, Stahl 2021
38 Mittelstadt 2019; Diese Debatte wird vertieft behandelt in Kempkens 2025.

triere, befasse sich die zweite Welle mit den Risiken, die von der derzeit bestehenden KI-Technologie ausgehe – dies sei eine Verlagerung vom Spekulativen zum Konkreten. Die Autorinnen verstehen die zweite Welle der KI-Ethik als Ausdruck einer grundlegend anderen Einstellung zur Frage, wo die relevanten Probleme liegen. Und sie identifizieren ein weiteres Problem, das aus ihrer Sicht durch die zweite Welle der KI-Ethik nicht abgedeckt wird: Nachhaltigkeit. Ihrer Meinung nach ist eine neue, dritte Welle der KI-Ethik erforderlich, um die mit der Entwicklung von KI-Systemen verbundenen Umweltkosten zu berücksichtigen. Während das Ziel der zweiten Welle die konkrete Analyse bestimmter KI-Systeme ist, geht es bei der dritten Welle darum, alle ethischen Bedenken zu berücksichtigen, die sich aus einer strukturellen Perspektive auf die Technologie ergeben – hier findet eine Verlagerung von einem isolationistischen zu einem strukturellen Ansatz statt.[39]

Bolte und van Wynsberghe werfen dem Ansatz der zweiten Welle Tech-Solutionismus (techno-solutionism) vor, weil die Autor:innen der zweiten Welle jede KI-Anwendung isoliert betrachten und annehmen, jedes Problem könne innerhalb des technischen Systems, das dieses Problem verursacht hat, gelöst werden. Die Erwartung ist dabei, dass die Gestaltung der KI-Systeme einen großen Einfluss auf ihre moralischen Eigenschaften und Wirkungen hat.[40] Doch laut Bolte und van Wynsberghe dient diese Kombination einer isolationistischen Sichtweise auf KI und einer techno-solutionistischen Ideologie nur den Interessen mächtiger Akteure.[41] Die dritte Welle hingegen betrachtet nicht nur Probleme einzelner KI-Anwendungen, sondern strukturelle Probleme, die durch den Einsatz von KI entstehen oder weiter bestehen.[42] Diese strukturellen Probleme werden hierbei den Autorinnen zufolge nicht auf technologische Rätsel reduziert: „The second wave approach relies on better design to bring about systemic or structural change. The third wave approach relies on systemic or structural change to bring about better design."[43] Strukturelle Probleme erscheinen in Ansätzen der zweiten Welle als

39 Bolte, van Wynsberghe 2024:1735f.
40 Ebd.:1739
41 Ebd.:1738
42 Ebd.:1740
43 Ebd.:1738

solche, die in der Technologie und durch technologisches Design korrigiert werden müssen. In den Ansätzen der dritten Welle hingegen werden Technologie und Technologiedesign als selbst in Strukturen und Systeme eingebettet konzipiert.[44]

Die vorliegende Arbeit bewegt sich zwischen der von Bolte und van Wynsberghe sogenannten zweiten und dritten Welle der KI-Ethik. Denn einerseits befasse ich mich konkret mit den Risiken, die derzeit von der bestehenden KI-Technologie ausgehen (zweite Welle), aber andererseits identifiziere ich diese Risiken mit Hilfe eines strukturellen Ansatzes (dritte Welle). Ich gehe nicht davon aus, dass die strukturellen ethischen Probleme innerhalb von einzelnen KI-Anwendungen gelöst werden können (Techno-solutionismus; zweite Welle), aber ich nehme auch nicht an, dass sich erst die Strukturen ändern müssen, bevor Technologie ethischer gestaltet werden kann (dritte Welle). Ich verfolge einen strukturellen Ansatz, der ethische Fragen auf einer systemischen Ebene aufdeckt (dritte Welle) und setze auf besseres Design, um strukturelle Veränderungen herbeizuführen (zweite Welle).

In Bezug auf den Einsatz von KI ist die größte Sorge in der Debatte der Digitalethik, dass Ungerechtigkeiten reproduziert werden.[45] Durch das Analysieren von Daten können Regelmäßigkeiten erkannt werden. Dies kann zum Beispiel dann nützlich sein, wenn man herausfinden will, wie viele der Mitarbeitenden bei Amazon weiblich sind. Dieses Beispiel zeigt, dass Daten nicht neutral sind – sie spiegeln gesellschaftliche Vorurteile und auch gefestigte Ungerechtigkeiten wider. Denn die Daten wurden in der Gesellschaft gesammelt, und in dem Maße, in dem die Gesellschaft Ungleichheit, Ausgrenzung oder andere Spuren von Diskriminierung aufweist, werden dies auch die Daten tun.[46] Durch eine reine Analyse können die Verhältnisse beschrieben und Ungerechtigkeiten reflektiert werden. Wenn die Daten aber als Trainingsdaten für Algorithmen verwendet werden, dann werden auch die Ungerechtigkeiten zur Entscheidungsgrundlage. Deshalb finden sich solche Vorurteile und Ungerechtigkeiten den Technologiephilosoph:innen Solon Barocas

44 Bolte, van Wynsberghe 2024:1739

45 Barocas, Selbst 2016:671; Goodman, Flaxman 2017:53; Whittlestone et al. 2019:198

46 Goodman, Flaxman 2017:53

und Helen Nissenbaum zufolge auch in algorithmischen Ergebnissen wieder: „(...) an algorithm is only as good as the data it works with."[47] Das Problem ist, dass algorithmische Ergebnisse oft als neutral wahrgenommen werden – das sind sie aber nicht. Denn ein Algorithmus kann nur wiedergeben, was in den Trainingsdaten steckt.

In der digitalethischen Debatte geht es also in Bezug auf den Einsatz von KI meist darum, ob KI in bestimmten Kontexten eingesetzt werden sollte und insbesondere darum, wer dann die Verantwortung dafür trägt. Denn Algorithmen des maschinellen Lernens, ein Teilbereich der KI, können erstaunliche Ergebnisse liefern und zum Beispiel, wie im Fall von ChatGPT, Aufsätze oder Code schreiben. Dabei sind ihre Ergebnisse jedoch unvorhersehbar.[48] Dadurch, dass solchen komplexen Algorithmen nicht die Entscheidungskriterien vorgegeben werden, sondern die KI sich diese selber sucht auf Basis sehr vieler bereits analysierter Datensätze, ist es für Menschen sehr schwer, ihre Ergebnisse, aber auch ihre Fehler zu verstehen, zu erklären oder vorherzusagen. Dem Rechtsberater Andrew Tutt zufolge macht das die KI sowohl wertvoll, als auch auch besonders gefährlich.[49] Denn die überraschenden Ergebnisse entstehen nur, weil der Algorithmus aus den ihm zur Verfügung gestellten Daten lernt, doch die Ergebnisse können falsch sein und enthalten sicher aus den Daten übernommene Vorurteile. Wer aber steht dafür gerade? In der Literatur ist die Rede von einer Verantwortungslücke[50], welche in Kapitel 7 genauer betrachtet wird.

Es gibt also viele ethische Probleme im Kontext der Digitalethik. Dem Technologiephilosoph Vincent C. Müller zufolge ist es wichtig diese früh zu erkennen: „The ethics of the design and use of computers is clearly an area of very high societal importance and we would do well to catch problems early on (...)."[51] Sobald Menschen nicht wissen, welche Handlung richtig wäre, gibt es ein ethisches Problem, das Aufmerksamkeit verdient. Oft handelt es sich dabei um in der Disziplin der Philosophie schon bekannte Probleme wie

47 Barocas, Selbst 2016:671
48 Tutt 2017:87
49 Ebd.:90
50 Matthias 2004; Sparrow 2007; Nyholm 2018; Himmelreich 2019; Kiener 2022:579; Oimann, Tollon 2024; Nyholm 2024
51 Müller 2023:15

zum Beispiel die Verletzung der Privatsphäre.[52] Diese hat auch schon vor der Digitalisierung aller Lebensbereiche eine Rolle gespielt und philosophische Konflikte hervorgerufen, bekommt aber durch die Digitalisierung eine verstärkte Bedeutung, weil das Offenbaren von persönlichen Informationen nun viel weitreichender ist. So sind zwar die Phänomene, mit denen sich die Digitalethik befasst, neu, aber die Fragen, die sie aufwerfen, sind alt und womöglich zeitlos. Laut Öhman und Watson ist es die Aufgabe der Philosophie, ihre grundlegenden Konzepte, die immer wieder zur Entschlüsselung neuer Phänomene beitragen, im Lichte neuer Herausforderungen, wie der Digitalisierung, neu zu definieren.[53]

Die Digitalethiker:innen Andréa Belliger und David Krieger versuchen nicht nur Konzepte neu zu definieren, sondern die digitale Ethik in ein neues Licht zu rücken und fordern eine Revolution der Ethik, die mit der Revolution der Technologie vergleichbar ist. Sie sind davon überzeugt, dass die Aufgabe der Ethik nicht darin besteht, Werte und Normen abzuleiten aus religiösen Lehren, Grundrechten, Erklärungen von Ethikkommissionen oder staatlichen Vorschriften und Richtlinien.[54] Solch ein Vorgehen ist im Kontext der Digitalethikdebatte jedoch tatsächlich zu beobachten. So haben zum Beispiel die Technologiephilosophen und Digitalethiker Luciano Floridi und Josh Cowls eine vergleichende Analyse von sechs Veröffentlichungen über Künstliche Intelligenz angestellt, um aus den Übereinstimmungen einen einheitlichen Rahmen zusammenzustellen. Die Veröffentlichungen stammen von einem Institut aus Belgien und den USA, einer kanadischen Universität, der Europäischen Kommission, dem Oberhaus des britischen Parlaments, einer Multistakeholder-Organisation in den USA und einer globalen Abhandlung nach dem Crowd-Sourcing-Prinzip mit Beiträgen von 250 globalen Vordenker:innen. Alle sechs Veröffentlichungen seien sehr renommiert und unmittelbar relevant für KI und ihre Auswirkungen auf die Gesellschaft als Ganzes. Floridi und Cowls haben darin insgesamt 47 Prinzipien identifiziert und diese in fünf Prinzipien zusammengefasst: Die vier biomedizinischen Prinzipien Achtung der Autonomie, Wohltätigkeit, Nichtschadensverursachung und Ge-

52 Müller 2023:16
53 Öhman, Watson 2019:5
54 Belliger, Krieger 2023:41

rechtigkeit sowie das Prinzip der Erklärbarkeit.[55] Dieses Vorgehen offenbart zwar einen erstaunlichen Konsens unter vornehmlich politischen und akademischen Autor:innen aus westlichen liberalen Demokratien, eine ethische Begründung für die Auswahl der Prinzipien gibt es jedoch nicht.

Digitalethik ist somit einerseits nicht dazu da, Werte und Normen aus Richtlinien abzuleiten, und andererseits ist sie auch keine philosophische Rechtfertigung für von oben verordnete Richtlinien, Gesetze und Regierungsvorschriften.[56] Laut Belliger und Krieger besteht digitale Ethik vielmehr aus konkreten Governance-Verfahren. Dabei bezeichnet Governance eine dezentrale, von unten nach oben gerichtete Regulierung von Netzwerken, welche als Designprozess geplant und umgesetzt wird.[57] Diese Netzwerke definieren sich, wie Datenökosysteme, über die Verbindung zwischen verschiedenen Akteuren mit der Fähigkeit, Informationen zu verarbeiten. Ihre Lösungsidee ist, dass jedes Netzwerk einen eigenen Governance-Rahmen entwickelt, in dem alle von dem Netzwerk Betroffenen (Stakeholder) sich an der Gestaltung beteiligen und sicherstellen, dass der Rahmen den Werten und Normen des Netzwerks entspricht. Die Werte und Normen des Netzes seien dabei Konnektivität, Fluss, Beteiligung, Transparenz, Authentizität und Flexibilität. Diese sechs Normen haben Krieger und Belliger in einer Veröffentlichung von 2014 aus Studien über sogenannte „neue Medien" (gemeint sind digitale Medien) abgeleitet, welche zeigen sollten, wie Netzwerken nach der digitalen Medienrevolution funktioniert. Die Autor:innen sind der Meinung, dass sie mit den sechs Normen die Netzwerknormen identifiziert haben, welche das kommunikative Handeln bestimmen.[58] „Since networks are dynamic processes, they follow certain inherent norms that describe what 'good' networking is. These values do not belong to a realm of what 'ought' to be as opposed to what 'is'. They describe, not prescribe."[59] Belliger und Krieger meinen, die sechs Normen leiten sich von den Möglichkeiten digitaler Technologien ab und beschreiben, was Netzwerke von

55 Floridi, Cowls 2021:8–9
56 Belliger, Krieger 2023:41, 52
57 Ebd.:40, 53
58 Krieger, Belliger 2014:10
59 Belliger, Krieger 2023:42

Natur aus für wünschenswert, angemessen und nützlich halten und was für ein gutes Funktionieren erforderlich ist.[60] Ihnen geht es hier nicht darum vorzuschreiben, wie digitale Netzwerke sein sollten, sondern nur darum zu beschreiben, wie sie sind. Die Werte und Normen dieser neuen digitalen Ethik sind also nicht mehr inhaltlich (substantive), sondern verfahrensbezogen (procedural): „(…) it is no longer *what* is held to be good by some person or group, which guides and legitimates action, but *how* anything at all is done well that becomes the foundation of ethics."[61] Die Grundlage der Digitalethik soll nach Belliger und Krieger demnach sein, wie ein Netzwerk gut funktioniert.

Hier liegt jedoch ein Missverständnis der Rolle der Ethik vor. Laut dem Philosophen Immanuel Kant beantwortet die Ethik, als Teilbereich der Philosophie, die Frage „Was soll ich tun?".[62] Die Antwort darauf ist nicht „Du solltest das Netzwerk so gestalten, dass es gut funktioniert". An dieser Stelle fangen die ethischen Fragen erst an: Für wen funktioniert das Netzwerk gut? Funktioniert es für alle Beteiligten gleichermaßen gut oder besser für diejenigen Akteure mit mehr Kontrolle? Die Aufgabe der Ethik ist es, zu hinterfragen was „gut" generell und in diesem Kontext bedeutet. Die reine Beschreibung eines Netzwerks und seiner Funktionsweisen findet auch in der Disziplin der Soziologie einen Platz. Aber die Ethik geht darüber hinaus und fragt, ob es das Netzwerk überhaupt geben sollte oder wie die einzelnen Beziehungen zwischen Akteuren innerhalb des Netzwerks gestaltet sein sollten, sodass es als ethisch bezeichnet werden kann. Wie man in dieser Arbeit sehen wird, gibt es unterschiedliche Ansätze bezüglich der Frage, wann etwas als ethisch bezeichnet werden kann. Konkret werden der Utilitarismus, die Tugendethik und Kants Deontologie dazu herangezogen.

Die Ethik spielt als reflektierende und hinterfragende Disziplin eine wichtige Rolle in der Debatte zur Digitalisierung – sie kann nicht einfach ersetzt werden durch soziologische Beschreibungen von Netzwerken oder durch rechtliche Festschreibungen. Das soll nicht bedeuten, dass die rechtlichen Entwicklungen in diesem Bereich nicht hilfreich oder notwendig sind. Im Gegenteil: Insbesonde-

60 Belliger, Krieger 2023:43
61 Ebd.
62 Kant 1956:727f.

re die neusten europäischen Verordnungen, die Datenschutzgrundverordnung (DSGVO) und die Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (AI Act), sind Meilensteine in der Regulierung des Umgangs mit Datenverarbeitung. Die DSGVO reguliert einerseits den „Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten“[63] und andererseits den „freien Verkehr solcher Daten“[64]. Die Verarbeitung personenbezogener Daten[65] ist nach DSGVO dann rechtmäßig, wenn die betroffene Person ihre Einwilligung für einen oder mehrere bestimmte Zwecke gibt; aber auch ein Vertrag, eine rechtliche Verpflichtung zum Schutz lebenswichtiger Interessen von natürlichen Personen, öffentliches Interesse und die Wahrung berechtigter Interessen von natürlichen Personen können rechtmäßige Gründe für die Verarbeitung personenbezogener Daten sein.[66] Dieser Datenschutz gilt nicht für anonyme Informationen, das heißt für Daten, bei denen es nur mit einem hohen zeitlichen und finanziellen Aufwand möglich wäre, die betroffenen Personen zu identifizieren.[67]

Der AI Act ist dazu da „(...) die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau (...) vor schädlichen Auswirkungen von KI-Systemen in der Union zu gewährleisten (...).“[68] Zu diesem Zweck verbietet der AI Act bestimmte Praktiken, wie zum Beispiel die Nutzung eines KI-Systems zum unterschwelligen Beeinflussen einer Person, zum Ausnutzen von Vulnerabilität oder Schutzbedürftigkeit von Personen, oder zum Bewerten von Personen mit dem Ergebnis ihrer Schlechterstellung oder Benachteiligung.[69] Im AI Act werden KI-Systeme anhand der von ihnen verursachten Risiken unterschieden. Sogenannte Hochrisiko-KI-Systeme dürfen nur dann eingesetzt werden, wenn es ein Risikomanagementsystem gibt, die Daten vorgeschriebenen Qualitätskriterien entsprechen, es eine technische Dokumentation gibt,

63 Art. 1, Abs. 1 DSGVO
64 Ebd.
65 Personenbezogene Daten sind laut DSGVO alle Daten, die eine Aussage über eine bestimmte Person machen (Art. 4, Abs. 1 DSGVO).
66 Art. 6, Abs. 1 a-f) DSGVO
67 Erwägungsgrund 26 DSGVO
68 Art. 1, Abs. 1 EU AI Act
69 Art. 5, Abs. 1 a)-c) EU AI Act

Aufzeichnungs- sowie Transparenzpflichten erfüllt sind, menschliche Aufsicht möglich ist und die KI-Systeme ein angemessenes Maß an Genauigkeit, Robustheit und Cybersicherheit aufweisen.[70] Als Hochrisiko-KI-Systeme gelten die Systeme, die in den folgenden Bereichen eingesetzt werden: Biometrie, kritische Infrastruktur, Bildung, Beschäftigung, Personalmanagement und Zugang zur Selbstständigkeit, Zugänglichkeit und Inanspruchnahme grundlegender privater und grundlegender öffentlicher Dienste und Leistungen, Strafverfolgung, Migration, Asyl und Grenzkontrolle, sowie Rechtspflege und demokratische Prozesse.[71]
Aber auch hier gibt es Ausnahmen. So sollen zum Beispiel KI-Systeme für Verwaltungsverfahren in Steuer- und Zollbehörden sowie für die Bekämpfung von Geldwäsche nicht als Hochrisiko-KI-Systeme eingestuft werden.[72] Denn der AI Act soll „(...) die Entwicklung und Verwendung innovativer Ansätze in der öffentlichen Verwaltung nicht behindern (...)"[73], solange das Risiko nicht zu hoch ist.

Solche rechtlichen Verordnungen, sind sehr wichtig, da man sich auf sie beziehen kann, wenn man Schutz verlangt bei der Datenverarbeitung und vor Schaden durch KI-Systeme. Aber die Ethik ist der Rechtsanwältin und Technologiephilosophin Mireille Hildebrandt zufolge vom Recht zu unterscheiden: „Ethics is both more and less than law: it is more because many ethical concerns are not addressed by the law and less because the outcome of ethical considerations are not necessarily transformed into legal norms and thus not enforceable by way of law."[74] Der Ethik fehlt die Durchsetzungskraft des Rechts, aber es ist auch nicht die Aufgabe der Ethik, bestimmte Regeln durchzusetzen. Im Gegenteil: Die Ethik sollte nicht abschließend in Regeln gegossen werden, da es verschiedene und sogar widersprüchliche ethische Haltungen geben kann, die gleichermaßen berechtigt sind. Ohne Einigkeit über die Ethik erfüllt erst mal das Recht mit Hilfe von festgelegten Regeln eine Rolle, die für Verlässlichkeit und Sicherheit sorgt.[75] Während das Recht die soziale

70 Art. 8 – 15 EU AI Act
71 Anhang III EU AI Act
72 Nr. 59 der nahestehenden Gründe (Preliminaries) EU AI Act
73 Nr. 58 der nahestehenden Gründe (Preliminaries) EU AI Act
74 Hildebrandt 2020:297
75 Ebd.:296f.

Ordnung aufrechterhalten soll, ist das Ziel der Ethik, die Menschen bei der Entscheidung anzuleiten, was über die Gesetze hinaus falsch ist und wie sie sich in bestimmten Situationen verhalten sollen. So schreibt der Elektroingenieur und Technologiephilosoph Spyros G. Tzafestas: „Laws provide a minimum set of standards for obtaining good human behavior. Ethics often provides standards that exceed the legal minimum."[76] Die Ethik geht über das Recht hinaus, weil viele ethische Bedenken durch Gesetze nicht berücksichtigt werden. Ethik kann an dieser Stelle Standards setzen, die nicht gesetzlich vorgeschrieben sind, und in Situationen, die nicht direkt vom Recht abgedeckt sind, eine Orientierungshilfe bieten.

So könnten laut der Technologiephilosophin und -ethikerin Esther Keymolen und der Professorin für internationale Datenverwaltung Linnet Taylor aus ethischen Gründen auch anonyme Daten als schützenswerte Daten behandelt werden – und zwar weil der Schutz der Privatsphäre für wichtig gehalten wird und nicht, weil man gesetzlich dazu verpflichtet ist: „In that case, they go beyond what the law expects from them and enter the realm of data ethics."[77] Datenethik beantwortet also Fragen danach, was getan werden sollte im Kontext von Datenerhebung und -verarbeitung. So lassen sich prinzipiell wichtige Aspekte herausarbeiten – wie in dieser Arbeit der Schutz der Privatsphäre und die Bereitstellung von Transparenz. Mit Hilfe von Gesetzen können rechtlich verpflichtende Mindeststandards für gutes menschliches Verhalten festgelegt werden. Da aber nicht alle ethischen Aspekte in Recht gegossen und sie oft durch Ausnahmen verwässert werden, geht das, was moralisch verpflichtend ist, darüber hinaus. Doch während der Verstoß gegen das Recht geahndet und mit Freiheits- oder Geldstrafen belegt werden kann, wird der Verstoß gegen moralische Verpflichtungen nur durch soziale Ächtung bestraft – dennoch sind sie wichtig.

Denn insbesondere im Bereich der Digitalisierung könnte das Recht laut Keymolen und Taylor durch schnelle technologische Entwicklungen an seine Grenzen stoßen: „For instance, developments in AI may lead to complex questions concerning explainability, fairness, and accountability which require us to rethink and reinterpret

76 Tzafestas 2018:98

77 Keymolen, Taylor 2023:493

existing legal norms."[78] Dieses Umdenken bestehender Rechtsnormen kann die Digitalethik leisten, weil sie viel beweglicher und flexibler ist als das Recht. Trotzdem kann die Ethik, auch laut Keymolen und Taylor, den Handlungsraum nicht alleine dominieren, da ihr die Durchsetzungskraft und Vorhersehbarkeit des Rechts fehlen. Denn wenn es keine rechtliche Regulierung gäbe, könnten Unternehmen ihre eigenen Spielregeln aufstellen, und andere Akteure müssten einfach mitspielen. Wenn allerdings Regeln aufgestellt und ohne Grundlage als „ethisch" bezeichnet werden, handelt es sich laut Keymolen und Taylor um Missbrauch der Digitalethik: „(...) in the darkest scenario, data ethics in the hands of data-driven businesses becomes a vehicle for intentional, malicious conduct, when it is used to hamper and ditch legal regulation."[79] Die Digitalethik ist nicht dafür da, kommerzielle Regeln zu rechtfertigen, sondern bietet die Möglichkeit zu durchdenken, was getan werden sollte. Digitalethik sollte also nicht dazu dienen, die Lücken zu füllen, welche das Recht nicht schnell genug füllen kann, sondern Digitalethik sollte im vom Gesetz vorgegebenen Handlungsspielraum dazu beitragen, einen verantwortungsvolleren Umgang mit der Digitalisierung zu entwickeln.[80] Auch Tzafestas plädiert für eine Balance: „(...) [F]or the best behavior, both law and ethics should be respected."[81] Denn im Rahmen des Rechts ist die Ethik dazu in der Lage, die Menschen auf unregulierte Probleme aufmerksam zu machen und Orientierung zu bieten.

Deshalb sollen in dieser Arbeit Datenökosysteme in der öffentlichen Verwaltung ethisch beleuchtet werden. Denn in Bezug auf die Digitalisierung von Verwaltungsdiensten kann die Ethik über das Recht hinaus Orientierung bieten. Das Onlinezugangsgesetz (OZG) legt fest, dass bei der Verbesserung des Onlinezugangs zu Verwaltungsleistungen Standards der IT-Sicherheit eingehalten werden sollen und Nutzerfreundlichkeit soll gegeben sein.[82] Bei Fragen bezüglich der Verarbeitung von personenbezogenen Daten bezieht sich das

78 Keymolen, Taylor 2023:494
79 Ebd.:495
80 Ebd.:494f.
81 Tzafestas 2018:98
82 §5 und §7 OZG

OZG auf die DSGVO.[83] Das OZG schafft aber eine weitere Ausnahme für Artikel 9, Absatz 1 der DSGVO, da die Verarbeitung von besonders sensiblen Daten erlaubt wird, wenn dies für Verwaltungsverfahren erforderlich ist.[84] Gleichzeitig schreibt das OZG die Einrichtung eines „Datenschutzcockpits" vor, mit dessen Hilfe sich die Bürger:innen über sie betreffende Datenübermittlungen zwischen öffentlichen Stellen informieren können.[85] Architekturvorgaben, Qualitätsanforderungen und Interoperabilitätsstandards müssen noch bis 2026 vom Bundesministerium des Innern und für Heimat zusammen mit dem IT-Planungsrat festgelegt werden.[86] Noch bevor rechtlich geregelt ist, wie genau der Onlinezugang zu Verwaltungsleistungen gestaltet wird, kann mit Hilfe der Ethik erörtert werden, wie diese Digitalisierung der öffentlichen Verwaltung, und damit die einzelnen Datenökosysteme gestaltet werden sollte. Im vom Gesetz vorgegebenen Handlungsspielraum kann die Ethik auf entscheidende Kriterien aufmerksam machen, die bei der Gestaltung von Datenökosystemen erfüllt sein sollten.

Aus diesem Grund möchte ich mit dieser Arbeit die digitalethische Debatte ergänzen um einen umfassenden Blick auf Datenökosysteme und wie diese auf Grundlage ihrer Risiken ethisch gestaltet werden können. Dabei ist es mir besonders wichtig, eine passende ethische Methode zu nutzen. In der Literatur herrscht jedoch Uneinigkeit darüber, welche ethische Methode im Kontext unbekannter, sich schnell entwickelnder Technologie verwendet werden sollte.[87] Um eine ethische Methode zu finden, die Orientierung bieten kann für die Gestaltung von Datenökosystemen, werde ich in dieser Arbeit zunächst einmal grundlegend analysieren, wie sich Datenökosysteme und die Risiken, die durch den Umgang mit digitalen Daten entstehen, ethisch bewerten lassen. Anschließend wird auf dieser Grundlage der ethischen Bewertung begründet, wie Datenökosysteme ethisch gestaltet werden können. Denn das *Argumentationsziel* dieser Arbeit ist es zu zeigen, wie der Umgang mit digitalen Daten in Datenökosystemen ethisch gestaltet werden kann. Zu diesem Zweck werden folgende Forschungsfragen erörtert:

83 §3a, Abs. 2 OZG; §8, Abs. 9 und 10 OZG
84 §8a, Abs. 1, Satz 2 OZG
85 §10, Abs. 1, Satz 1 OZG
86 §6, Abs. 1 OZG
87 Vallor 2016:7f.

1. Wie lassen sich Datenökosysteme ethisch bewerten?
2. Wie sollten Datenökosysteme dementsprechend gestaltet werden?

Um herausfinden zu können, wie sich Datenökosysteme ethisch bewerten lassen und somit die erste Forschungsfrage beantworten zu können, wird zuerst eine Konzeption von Datenökosystemen aufgestellt (Kapitel 2). Daraus wird in Kapitel 3 ersichtlich werden, dass im Datenökosystem zwischen den Akteuren das Wissen und die Möglichkeiten zur Kontrolle ungleich verteilt sind, und die Akteure mit mehr Kontrolle den Akteuren mit weniger Kontrolle zwei Risiken auferlegen: das Risiko der Verletzung der Privatsphäre und das Risiko der fehlenden Transparenz. Die Methode zur ethischen Bewertung von Datenökosystemen muss deshalb sowohl die ungleichen Wissensstände und Kontrollmöglichkeiten als auch die daraus erwachsende Auferlegung der Risiken bewerten können. Es wird deutlich, dass Utilitarismus und Tugendethik diese Bedingungen nicht erfüllen und die Deontologie die beste Grundlage für die ethische Bewertung von Datenökosystemen darstellt (Kapitel 4). Was ethisch bewertet werden muss, ist die Struktur von Datenökosystemen, sich durch ungleich verteiltes Wissen, ungleiche Kontrollmöglichkeiten der Akteure sowie dadurch mögliche Risikoauferlegung auszeichnet.

Deshalb wird in Kapitel 5 die Frage beantwortet, unter welchen Bedingungen es zulässig ist, anderen ein Risiko aufzuerlegen. Zur Bestimmung dieser Bedingungen wird eine deontologische Risikoethik entwickelt, indem die Risikoethik um die dritte Formel des Kategorischen Imperativ von Kant ergänzt wird. Es wird deutlich, dass es nur dann zulässig ist, anderen Personen ein Risiko aufzuerlegen, wenn diese dabei nicht als bloßes Mittel gebraucht werden. Deshalb darf ein Risiko einer Person nur nach deren informierter Zustimmung auferlegt werden. Datenökosysteme lassen sich also ethisch bewerten mit der von mir auf der Grundlage von Kants Deontologie entwickelten deontologischen Risikoethik.

Um herausfinden zu können, wie Datenökosysteme dementsprechend gestaltet werden sollten, und damit die zweite Forschungsfrage beantworten zu können, werden in Kapitel 6 und 7 die Risiken genauer betrachtet. Denn es ist nicht nur die Risikoauferlegung selbst ethisch relevant, sondern auch die weitestgehende Vermeidung der Risiken ist geboten. Dies wird gezeigt durch die Begründung einer Pflicht zum Schutz der Privatsphäre (6.3) sowie einer

Pflicht zur Bereitstellung transparenter Ergebnisse (7.3). Die beiden Risiken werden einer gründlichen Betrachtung unterzogen, um zu entwickeln, wie sich Privatsphäre und Transparenz bei der Datenverarbeitung in einem Datenökosystem schützen und fördern lassen (Kapitel 6 und 7). Zum Schluss werden mit Hilfe der deontologischen Risikoethik administrative Anforderungen abgeleitet, die bei ihrer Einhaltung den Schutz und die Förderung von Privatsphäre und Transparenz ermöglichen.

Daraufhin stellt sich in Kapitel 8 die Frage nach der Verantwortung, denn wenn verantwortungsvolles Handeln unterbleibt, ist es möglich, dass die Privatsphäre der Datenquellen oder die Transparenz für die Datennutzenden verletzt werden. Die Verantwortung ist der wahrscheinlich stärkste Durchsetzungsmechanismus der Ethik und kann denjenigen Akteuren zugeschrieben werden, die Kontrolle über das zu Verantwortende haben. Aus einer Position der Kontrolle über den Umgang mit Daten innerhalb eines Datenökosystem erwächst also Verantwortung für die Minimierung der Risiken und das ethische Gestalten.

Auf der Suche nach einer Umsetzungsmethode, mit der die ethischen Erkenntnisse wirksam in die Praxis überführt werden können, leite ich aus der bisherigen Argumentation Kriterien für eine ethische Gestaltung ab. Es liegen bereits verschiedene Vorschläge zur Umsetzung digitaler Ethik in die Praxis vor, die ich einteile in ethische Leitlinien, ethische Gestaltungsprozesse und Risikoanalysen. Jeder dieser Ansätze weist Lücken auf, wie ich in Kapitel 9 ausführlich aufzeige. Deshalb entwickele ich abschließend einen eigenen Verfahrensmaßstab zur Umsetzung digitaler Ethik in die Praxis, um die Vorteile von Datenökosystemen gefahrlos nutzen zu können. Als Antwort auf meine zweite Forschungsfrage stelle ich in Kapitel 10 meinen ethischen Verfahrensmaßstab vor, welcher expliziert, wie Datenökosysteme den Erkenntnissen dieser Arbeit zufolge gestaltet werden sollten. Folglich wird abschließend gezeigt, wie der Umgang mit digitalen Daten in Datenökosystemen ethisch gestaltet werden kann.

Beginnen wir das Vorhaben mit einer ausführlichen Betrachtung des zu bewertenden und gestaltenden Gegenstands: dem Umgang mit Daten im Datenökosystem.

2 Umgang mit Daten in Datenökosystemen

Dieser erste Baustein meiner Argumentation schafft die Grundlage für das Vorhaben zu zeigen, wie der Umgang mit digitalen Daten in Datenökosystemen ethisch gestaltet werden kann. Dafür wird zuerst einmal definiert was Daten sind und wie diese mit Hilfe von Algorithmen verarbeitet werden können. Dann definiere ich Datenökosysteme über den Datenfluss zwischen vier Akteuren – Datenquellen, Datengenerierenden, Datenverarbeitenden und Datennutzenden – mit verschiedenen Rollen und stelle ein eigenes Modell von der Struktur eines Datenökosystems auf. Dabei zeigt sich, dass das Sammeln und Verarbeiten von Daten sowohl Vorteile hat, als auch Risiken birgt. Dieser für meine Argumentation zentrale Begriff des Risikos wird definiert und ich zeige, dass im Datenökosystem nicht alle Akteure über das gleiche Wissen verfügen bezüglich der Datenverarbeitung im Datenökosystem und dass ungleiche Kontrollmöglichkeiten vorherrschen, weshalb auch die Risiken ungleich verteilt sind.

2.1 Datenverarbeitung

Um sich mit der ethischen Bewertung und Gestaltung von Datenverarbeitung in Datenökosystemen auseinandersetzen zu können, muss zuerst einmal geklärt werden, was genau mit Daten gemeint ist. Die Technologieforscher Syed Iftikhar Hussain Shah, Vasilis Peristeras und Ioannis Magnisalis definieren Daten folgendermaßen: „Data are facts and figures about an object (...)."[88] Die Daten, von denen in dieser Arbeit die Rede ist, sind digital gespeichert und sie liegen ursprünglich in roher Form vor, zum Beispiel als Dokumente, Transkripte oder audio-visuelle Aufnahmen.[89] Die Daten können dabei

88 Shah et al. 2020:102
89 Poikola et al. 2011:13

über alles Mögliche eine Aussage machen, zum Beispiel über die Umwelt, Gegenstände oder Personen. In dieser Arbeit sind vor allem die Daten von Interesse, die eine Aussage über Personen machen – ich nenne sie personenbezogene Daten. Welche Daten personenbezogen sind, ist umstritten, aber in dieser Arbeit wird in Übereinstimmung mit der Datenschutzgrundverordnung (DSGVO) angenommen, dass alle Daten, die eine Aussage über eine bestimmte Person machen, personenbezogene Daten sind.[90] Personenbezogene Daten offenbaren also direkt oder indirekt die Identität von bestimmten Personen – hier kann es sich um den Namen, die Anschrift oder auch um Fotos der Person handeln.[91]

Personenbezogene Daten können dann als private Daten bezeichnet werden, wenn die Personen, über welche die Daten eine Aussage machen, diese Daten nicht der Öffentlichkeit preisgeben möchten. Eine Teilmenge der privaten Daten wird als sensible Daten bezeichnet, wenn die betroffenen Personen sie nicht öffentlich machen wollen, weil sie bei einer Offenbarung mit negativen Auswirkungen rechnen. Dies betrifft Daten, die Auskünfte enthalten über ethnische Herkunft, politische Meinungen, religiöse Überzeugungen, Gesundheit oder sexuelle Orientierung einer bestimmten Person.[92] Dabei gelten solche Daten als direkte Identifikatoren, die eine Person unmittelbar erkennbar machen, wie zum Beispiel Name, Anschrift oder Fotos. Als indirekte Identifikatoren werden solche Daten bezeichnet, die nicht direkt auf die Identität einer Person hinweisen, aber durch die Kombination mit anderen Daten auf eine Person schließen lassen. So können zum Beispiel auch die Postleitzahl und das Alter indirekte Identifikatoren sein, wenn zum Beispiel in einem Ort nur eine 90-jährige Frau wohnt.[93]

Um große Mengen von Daten verstehen oder interpretieren zu können, müssen sie analysiert werden. Damit rohe Daten zur Verarbeitung bereit sind, werden sie homogenisiert und in ein maschinenlesbares Format gebracht.[94] Eine solche Analyse kann erleichtert werden, durch den Einsatz von Algorithmen, die bestimmte Muster

90 Art. 4, Abs. 1, DSGVO
91 Zainab, Kechadi 2019:2
92 Art. 9, Abs. 1 DSGVO; Zainab, Kechadi 2019:2f.
93 Zainab, Kechadi 2019:3
94 Poikola et al. 2011:29, 32

in den Daten erkennen. So kann durch eine algorithmische Analyse aus Daten Wissen gewonnen werden.[95] Denn die Datenverarbeitung erzeugt neue Informationen, die nicht mehr als Rohdaten bezeichnet werden können.[96] Wenn sehr große Datenmengen analysiert werden sollen, benötigt man neue Methoden der Datenverarbeitung – bei solch großen Mengen von Daten spricht man von Big Data[97], und der Prozess der Wissensgewinnung, -darstellung und -anwendung nennt sich Data Mining.[98] Durch das Erkennen von Mustern in diesen vielen Datensätzen wird es möglich, Aussagen über ihren Inhalt zu treffen, die wirtschaftlich, sozial, technisch oder rechtlich relevant sein können. Der Glaube, dass große Mengen an Datensätzen zu Einsichten führen können, die zuvor unmöglich waren, ist laut Medienwissenschaftlerin und Sozialforscherin Danah Boyd sowie KI-Forscherin Kate Crawford jedoch ein Mythos.[99] Die bereits erwähnten Technologiephilosoph:innen Solon Barocas und Helen Nissenbaum aber halten dagegen, dass Erkenntnisse durch eine Analyse großer Datenmengen unerwartet sein können und dass darin der Wert der Daten liegt.[100]

Diese Analyse großer Datenmengen wird durch Algorithmen ermöglicht. Die Informatiker Thomas H. Cormen, Charles E. Leiserson, Ronald Rivest und Clifford Stein definieren Algorithmen wie folgt: „Somit ist ein Algorithmus eine Folge von Rechenschritten, die die Eingabe in die Ausgabe umwandeln."[101] Es handelt sich bei einem Algorithmus um ein Hilfsmittel zur Lösung von Rechenproblemen. Zum Beispiel können Zahlen in eine andere Reihenfolge sortiert werden. In der Eingabe sind die Zahlen unsortiert, in der Ausgabe sollen sie aber in aufsteigender Reihenfolge erscheinen. Der Algorithmus enthält die Angaben, die notwendig sind, um die Zahlen in aufsteigender Reihenfolge zu sortieren. Dafür muss definiert sein, was mit „in aufsteigender Reihenfolge" gemeint ist. Cormen et al. fügen hinzu: „Ein Algorithmus wird als korrekt bezeichnet, wenn

95 Immonen et al. 2014:89
96 Poikola et al. 2011:74
97 Shah et al. 2020:102; Altintac 2022:5; Corsten, Roth 2022:4; Donges 2022:214
98 Ertel 2021:206
99 Boyd, Crawford 2012:663
100 Barocas, Nissenbaum 2014:60
101 Cormen et al. 2013:5

er für jede Eingabeinstanz mit der korrekten Ausgabe stoppt."[102] Ein korrekter Algorithmus löst das gegebene Rechenproblem und würde in diesem Fall die eingegebenen Zahlen in der definierten Weise in aufsteigender Reihenfolge sortieren. Ein Algorithmus ist also eine präzise und eindeutige Handlungsanweisung, welche dazu führt, dass der ihn nutzende Mensch aus der Eingabe die erwartete Ausgabe erhält.

So kann zum Beispiel mit Hilfe eines Algorithmus schnell ausgerechnet werden, ob Antragstellende ein Anrecht auf Wohngeld haben und in welcher Höhe ihnen dieses Geld nach dem Wohngeldgesetz (WoGG) zusteht. Dafür müssen die Bedingungen des Wohngeldgesetzes in den Algorithmus einprogrammiert werden. Es muss also konkret vorgegeben werden, inwiefern sich das Wohngeld nach der Anzahl der zu berücksichtigenden Haushaltsmitglieder, der zu berücksichtigenden Miete oder Belastung und dem Gesamteinkommen richtet und unter welchen Bedingungen man keinen Anspruch auf Wohngeld hat.[103] Dann kann dieser Algorithmus aus den Daten von Antragstellenden berechnen, ob ein Anspruch besteht und in welcher Höhe. In dieser Arbeit nenne ich solche Algorithmen Wenn-Dann-Algorithmen. Denn damit wird eine Bedingung klar vorgegeben, zum Beispiel *wenn* „das Wohngeld weniger als 10 Euro monatlich betragen würde"[104], *dann* besteht kein Wohngeldanspruch. Solche Wenn-Dann-Algorithmen können Rechenprobleme schneller lösen als Menschen und diese somit zum Beispiel bei der Bearbeitung von Wohngeldanträgen entlasten.

Die „Digitization" hat einen solchen Einsatz von Algorithmen zur Problemlösung ermöglicht. Während es bei der Digitalisierung zuerst nur um die Automatisierung von Prozessen ging, geht es mittlerweile um die Autonomisierung von Prozessen.[105] Das heißt, dass inzwischen technische Systeme selbstständig lernen und Lösungen anbieten sollen. Diese Systeme werden als Künstliche Intelligenz (KI) bezeichnet. Die Bezeichnung „Künstliche Intelligenz" stammt aus einem Antrag für ein Forschungsprojekt von den Informatikern und Mathematikern John McCarthy, Marvin Minsky, Nathaniel Roches-

102 Cormen et al. 2013:6
103 §4 – 21 WoGG
104 §21 WoGG
105 Klenk et al. 2020:5f.

ter und Claude Elwood Shannon im Jahr 1955. Damals schrieben sie: „The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."[106] Das Ziel war, menschliche Intelligenz von einer Maschine simulieren zu lassen. Diese Maschine sollte Probleme lösen, Konzepte formulieren, Sprache verwenden und sich selbst verbessern können. Einer der Forscher, Marvin Minsky, definiert KI so: „(…) *artificial intelligence*, the science of making machine do things that would require intelligence if done by men."[107] Das Ziel der Wissenschaftler war es, Maschinen dazu zu bringen, Dinge zu tun, die Intelligenz erfordern würden, wenn sie von Menschen ausgeführt würden.

Da die Verwendung des Intelligenzbegriffs uns in dessen kontroverse Definitionen hineinzieht, bietet es sich an, alternative Definitionen von KI zu betrachten. So zum Beispiel die Definition der Informatikerin Elaine A. Rich: „Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better."[108] Die Definition umgeht die inhaltliche Aufschlüsselung des Intelligenzbegriffs, schreibt Intelligenz aber grundsätzlich dem Menschen zu, und gleichzeitig hält er diesbezüglich die aktuelle Überlegenheit des Menschen gegenüber dem Computer fest. Konkreter formulieren dies die Informatiker Uwe Lämmel und Jürgen Cleve: Künstliche Intelligenz ist ein „Teilgebiet der Informatik, welches versucht, menschliche Vorgehensweisen der Problemlösung auf Computern nachzubilden, um auf diesem Wege neue oder effizientere Aufgabenlösungen zu erreichen."[109] Intelligenz wird hier gleichgesetzt mit menschlicher Problemlösung. Die Definition geht allerdings darüber hinaus und erwartet von der Künstlichen Intelligenz neue und effizientere Vorgehensweisen.

Daraufhin stellt sich die Frage, wozu die Künstliche Intelligenz fähig ist. Aktuell gibt es nur sogenannte schwache KI (Weak AI), das heißt Programme, die Fähigkeiten der Intelligenz demonstrieren, ohne selbst „intelligent" zu sein. Viel diskutiert wird auch die Entwicklung von starker KI (Strong AI), welche ein Programm sein

106 McCarthy et al. 1955:2
107 Minsky 1968:v (Preface)
108 Rich 1983; Vgl. Ertel 2021:2
109 Lämmel, Cleve 2008:14

soll, das ein Bewusstsein hat, was aktuell nicht existiert.[110] Die Frage, ob Computerprogramme irgendwann Bewusstsein erlangen können oder sollten, fällt unter die von Bolte und van Wynsberghe sogenannte erste Welle der KI-Ethik und soll hier nicht weiter behandelt werden.[111] In dieser Arbeit ist mit KI immer schwache KI gemeint, und ich gehe davon aus, dass Computerprogramme nur einem Regelsatz folgen, um Symbole abgleichen und Input in Output verwandeln zu können.

Wer den Regelsatz bestimmt, unterscheidet sich bei verschiedenen Typen von Algorithmen. Während bei Wenn-Dann-Algorithmen der Regelsatz klar vom Menschen vorgegeben wird, wird bei KI-Systemen[112] nur noch das Ziel des Outputs vorgegeben. Die dabei der Maschine vorgegebenen Werte oder Ziele sollen mit denen der Menschen übereinstimmen – da dies nicht immer einfach ist, nennt sich der Versuch, diese Übereinstimmung herzustellen, das Problem der Werteanpassung (value alignment).[113] Der Regelsatz für die Symbolverarbeitung kann also nicht nur vorgegeben werden, sondern er kann auch durch KI-Systeme erschlossen werden – diese Methode nennt sich Machine Learning (ML; Maschinelles Lernen). Hierbei handelt es sich um ein Teilgebiet der Künstlichen Intelligenz, welches statistische Methoden verwendet, um Maschinen in die Lage zu versetzen, sich durch Erfahrung zu verbessern.[114] Beim maschinellen Lernen liest ein Computer einige Daten, erstellt ein Modell auf der Grundlage dieser Daten und verwendet das Modell sowohl als These über die Welt als auch als Software, die Probleme lösen kann. Die Informatiker Stuart J. Russel und Peter Norvig definieren dieses Lernen so: „An agent is learning if it improves its performance after making observations about the world."[115] Dem liegt das Prinzip der

110 Searle 1980:417; Boddington 2023:3f.
111 Bolte, van Wynsberghe 2024:1735f
112 Hierbei handelt es sich nicht um ein soziotechnisches System, wie bei Datenökosystemen, sondern um ein rein technisches System. Das heißt, hier stehen verschiedene technische Objekte in bestimmten Beziehungen zueinander (Shapiro 1997:73).
113 Russel, Norvig 2021:22
114 Mitchell 1997:XV; Ertel 2021:201
115 Russel, Norvig 2021:669; Hervorhebung entfernt

Induktion zugrunde, da aus wiederkehrenden Mustern in Datensätzen allgemeine Regeln abgeleitet werden.[116]

Es gibt verschiedene Ansätze des Maschinellen Lernens, man unterscheidet supervised learning, unsupervised learning und reinforcement learning. Beim supervised learning (Überwachtes Lernen) beobachtet das KI-System Eingabe-Ausgabe-Paare und erlernt die Funktion, welche Eingabe und Ausgabe verbindet. Dabei werden die Eingabe-Ausgabe-Paare von den Programmierenden vorgegeben, zum Beispiel Fotos (Eingabe), die beschriftet werden mit den Dingen, die auf ihnen abgebildet sind, wie „Katze" oder „Hund" (Ausgabe). Solche Ausgaben werden als „Label" bezeichnet und ermöglichen den KI-Systemen bei neuen Fotos das passende Label vorherzusagen.

Beim unsupervised learning (Unüberwachtes Lernen) identifiziert das KI-System Muster in Datensätzen (Eingaben) und sortiert die Daten, die sich ähneln, in Gruppen (Ausgabe). So können zum Beispiel viele Bilder in Gruppen ähnlicher Bilder eingeteilt werden – entweder Bilder mit Katzen oder Bilder mit Hunden.

Beim reinforcement learning (Verstärkungslernen) lernt das KI-System durch das Feedback der Programmierenden. Dabei wird dem KI-System zurückgemeldet, ob seine Ausgabe richtig oder falsch war: „Ja, das Bild zeigt eine Katze" oder „Nein, das Bild zeigt keine Katze". Daraufhin verändert das KI-System sein Regelsystem, um möglichst viele richtige Ausgaben generieren zu können.[117]

Maschinelles Lernen ist hilfreich, wenn die Programmierenden nicht alle möglichen zukünftigen Situationen vorhersehen können oder wenn sie nicht wissen, wie sie selbst eine Lösung programmieren sollen. In solchen Fällen gibt es für die Maschine verschiedene Lernprobleme: Klassifizierung und Regression. Wenn der Output ein endlicher Satz von Werten sein soll, nennt sich das Klassifizierung, und wenn der Output eine Zahl sein soll, nennt sich das Regression.[118] Das geschieht mit Hilfe von verschiedenen mathematischen Modellen, wie zum Beispiel Entscheidungsbäumen oder linearen Regressionen. Je komplexer die Aufgabe für die KI-Systeme sind, desto komplexer werden auch die mathematischen Modelle,

116 Hume 2000[1739]; Russel, Norvig 2021:24, 670; Boddington 2023:3
117 Russel, Norvig 2021:671
118 Ebd.:669f.

welche das Erlernen eines passenden Regelsatzes ermöglichen sollen. Hier gibt es zum Beispiel Random Forests (Zufallswälder) oder Neuronale Netze, die so komplex sind, dass die Entstehung ihrer Regelsätze und auch die Regelsätze selbst für die Programmierenden nicht mehr nachvollziehbar sind.[119]

Für das Programmieren von Neuronalen Netzen wird eine hochkomplexe Technik verwendet: das Deep Learning, ein Teilgebiet des Machine Learning.[120] Das Deep Learning unterscheidet sich vom Machine Learning dadurch, dass die Berechnungswege von Eingabe zu Ausgabe viele, nicht nachvollziehbare Schritte umfassen. Hierdurch können auch unstrukturierte Daten wie Texte, Bilder, Töne und Videos verarbeitet werden.[121] Solche hochkomplexen KI-Systeme, die mit der Methode des Machine Learnings, oder auch konkreter mit Techniken des Deep Learnings, ihre eigenen Regelsysteme erstellen, sind nicht mehr interpretierbar. Russel und Norvig definieren Interpretierbarkeit folgendermaßen: „We say that a machine learning model is interpretable if you can inspect the actual model and understand why it got a particular answer for a given input, and how the answer would change when the input changes."[122] Wenn nicht mehr verständlich ist, warum das KI-System für eine bestimmte Eingabe eine bestimmte Ausgabe generiert und wie sich die Ausgabe ändern würde, wenn sich die Eingabe ändert, dann ist der erlernte Regelsatz unbekannt und das KI-System nicht interpretierbar. Will man trotzdem verstehen, warum das KI-System für eine bestimmte Eingabe eine bestimmte Ausgabe generiert und somit den erlernten Regelsatz durchschauen, muss man sich auf Methoden der Erklärbarkeit (Explainable AI; XAI) verlassen. Während eine Inspektion des eigentlichen KI-Systems Interpretierbarkeit ermöglicht, kann Erklärbarkeit nur durch einen separaten Prozess bereitgestellt werden, indem ein anderer Algorithmus zusammenfasst, wie das zu erklärende KI-System vorgeht.[123]

119 Russel, Norvig 2021:675–722
120 Ertel 2021:321; Russel, Norvig 2021:801ff.
121 Russel, Norvig 2021:801
122 Ebd.:729; Hervorhebung entfernt
123 Ebd.:729

Bei dieser autonomisierten Form der Datenverarbeitung mit Hilfe von KI-Systemen entstehen mehrere Risiken, die im Lauf der Arbeit vertieft erörtert werden. An dieser Stelle ist es wichtig festzuhalten, dass Wenn-Dann-Algorithmen bei gleichem Input immer den gleichen Output generieren – KI-Systeme hingegen können bei gleichem Input auch unterschiedlichen Output generieren, da sie mit jedem neuen Input ihr Regelwerk überarbeiten. Das deutet daraufhin, dass KI-Systeme nicht fehlerfrei sind – in Kombination mit der fehlenden Interpretierbarkeit von hochkomplexen Modellen wird das laut dem Informatiker Wolfgang Ertel zum Problem: „Die neue Qualität der Deep Learning Systeme(...) ist nun, dass es sehr schwer bis unmöglich wird, eine Stelle im System als Ursache des Fehlers zu identifizieren."[124] Das erschwert den Umgang mit fehlerhaften Ergebnissen erheblich und mahnt zur Vorsicht beim allgemeinen Umgang mit algorithmischen Ergebnissen. Denn eigentlich werden Fakten in Diskursen produziert, sodass ihre Entstehung am Ende nachvollziehbar ist.[125] Ein Algorithmus aber macht den Diskurs mit sich selbst aus und produziert nur eine Behauptung, deren Entstehung unter Umständen nicht nachvollziehbar ist.

Deshalb ist es wichtig, die Ergebnisse von Algorithmen nur als Sinnangebot zu verstehen, das noch eines Diskurses bedarf. Sie dürfen nicht als unstrittig verstanden werden, da es sich um gedeutete Daten handelt. Algorithmische Ergebnisse sind keine Tatsachen, da sie noch nicht gesellschaftlich akzeptiert werden konnten. Sie müssen sich als Sinnangebot noch einem gesellschaftlichen Diskurs stellen, wie jede menschliche Deutung von Daten auch. Wenn die Ausgaben eines Algorithmus als Fakten betrachtet werden, wird er zum Diskursakteur. Er darf dabei aber nicht mehr Macht haben als die Menschen, die nach der Entstehung dieser Fakten fragen. Der algorithmische Beitrag zum Diskurs kann gesellschaftliche Veränderungen anregen. Solch große Einflussmöglichkeiten erfordern die Übernahme von Verantwortung. Wer diese Verantwortung zu tragen hat, soll im Verlauf dieser Arbeit geklärt werden.

Eine Problemlösung für die Interpretierbarkeit der KI-Systeme und also die Identifikation von Fehlern können die Methoden der

124 Ertel 2021:343
125 Felder 2013:14ff.

Erklärbarkeit sein. So konnten zum Beispiel die Informatiker Marco Tulio Ribeiro, Sammer Singh und Carlos Guestrin zeigen, dass ein KI-System, welches Bilder von Huskys und von Wölfen unterscheiden kann, einen völlig falschen Regelsatz erlernt haben kann: „(...) the classifier predicts “Wolf" if there is snow (or light background at the bottom), and “Husky" otherwise, regardless of animal color, position, pose, etc.“[126] In diesem Fall wurde das KI-System absichtlich dazu gebracht, falsche Regeln zu erlernen, da die Trainingsdaten aus Bildern bestanden, welche nur dann Schnee im Hintergrund hatten, wenn Wölfe abgebildet waren, während Bilder von Huskys keinen Schnee im Hintergrund hatten. Aber solche Fehler können auch unabsichtlich geschehen und in Kontexten, in denen Menschen von den Ausgaben des KI-Systems betroffen sind, zu negativen Konsequenzen führen. So zum Beispiel, wenn eine Gesichtserkennung dunkelhäutige Frauen häufig falsch klassifiziert, hellhäutige Männer hingegen kaum. Die Informatikerinnen Joy Buolamwini und Timnit Gebru haben festgestellt, dass dies daran lag, dass die Trainingsdatensätze überwiegend aus Bildern von hellhäutigen Personen bestanden.[127] Auf dieser Grundlage hat das KI-System falsche Regeln erlernt.

Nicht immer ist so deutlich, warum das KI-System falsche Regeln erlernt hat oder nach welchen Regeln es überhaupt entscheidet. Eine einfache Erklärung dieser Regeln kann den bereits erwähnten Informatikern Stuart J. Russel und Peter Norvig zufolge zu einem falschen Gefühl der Sicherheit führen: „After all, we typically choose to use a machine learning model (rather than a hand-written traditional program) because the problem we are trying to solve is inherently complex, and we don’t know how to write a traditional program.”[128] Komplexe Probleme erfordern komplexe Lösungen und diese komplexen Lösungen lassen sich nicht unbedingt einfach erklären. Die selbsterlernten Regelsätze von KI-Systemen, die nicht interpretierbar sind, bleiben oft unverständlich. Trotzdem werden diese technischen Möglichkeiten verstärkt genutzt.[129]

126 Ribeiro et al. 2016:1142
127 Buolamwini, Gebru 2018
128 Russel, Norvig 2021:730
129 Ebd.

An dieser Stelle fragt sich, warum so viele Daten mit Hilfe von Wenn-Dann-Algorithmen und KI-Systemen verarbeitet werden. Shah et al. finden folgende Antwort darauf: „Data by itself has no value. However, the extraordinary value-adding potential of data lies in the ability to extract meaningful and actionable information from it."[130] Die algorithmische Analyse der Daten verspricht ein wertvolles Ergebnis. Wenn viele Daten kombiniert und unter ihnen Muster gefunden werden, kann dadurch neues Wissen entstehen. Wenn also zum Beispiel viele Daten über langjährige Pflanzenpflege gesammelt werden, machen diese in der Summe eine Aussage darüber, bei welcher Pflege welche Pflanze am längsten lebt. So entsteht Wissen darüber, welche Pflanzen viel Wasser brauchen und welche kein direktes Sonnenlicht vertragen – Daten über Pflanzenpflege versprechen Wissen darüber, wie Pflanzen am besten gepflegt werden können. So können Kausalitäten entdeckt werden, die zukünftig bessere Entscheidungen ermöglichen.

Da die Analyse von Daten sehr schnell zu Entscheidungsvorschlägen führen kann, sind digitale Daten zum wichtigsten Rohstoff geworden – sie werden sogar als das Öl des 21. Jahrhunderts bezeichnet.[131] Das Wertvolle an den Daten ist ihr Potenzial, durch ihre Interpretation neues Wissen zu erschaffen. Dieses neue Wissen ermöglicht dann Innovationen und damit auch finanzielle Vorteile. Wenn man zum Beispiel besser über die Pflege von Pflanzen Bescheid weiß, kann man den Ertrag pro Bodenfläche vergrößern und somit effektiver Landwirtschaft betreiben.

Dieses Wissen darf aber nicht unhinterfragt akzeptiert werden, weil Daten nicht neutral sind. Der Germanist Ekkehard Felder hingegen meint Daten seien „(...) unstrittige, allseits akzeptierte Fakten (...)."[132] So herrscht in der Literatur die Meinung vor, dass Daten in dem Sinne *gegeben* sind, dass sie überprüfbar sind, da sie durch Verhalten und Beobachtungen von Menschen entstanden sind. Distanzen sind messbar, Bewertungen von Schulleistungen sind nachweisbar etc. Jedes Datum für sich mag neutral sein, solange es überprüfbar ist und als deskriptive Aussage für sich steht. Auch Daten, die Aussagen über Menschen machen, sind erstmal deskriptiv. Aber

130 Shah et al. 2020:110
131 Streibich 2015:16
132 Felder 2013:14

wenn diese deskriptiven Aussagen für bare Münze genommen werden, kann es dadurch zu Diskriminierung kommen.

Diskriminieren bedeutet ursprünglich einfach nur unterscheiden (lat. discriminare), aber der Begriff wird seit dem 20. Jahrhundert mit einer negativen Wertung verbunden.[133] Dem Sozialpsychologen Andreas Zick zufolge lässt sich dies folgendermaßen erklären: „Diskriminierung beginnt bei der Wahrnehmung und endet in einer Herstellung und Etablierung von Ungleichwertigkeit, die Ungleichheit begründen soll."[134] Das heißt gemäß der Philosophin und Kommunikationswissenschaftlerin Larissa Krainer, dass Diskriminierung zwei verschiedene Prozesse beschreibt: Zum einen die Unterscheidung verschiedener Gruppen und zum anderen die unterschiedliche Bewertung dieser Gruppen. Letzteres funktioniert laut Krainer so: „Zur Abwertung von Gruppen werden häufig Stereotype entwickelt und angewandt."[135] Ein solches Stereotyp ist zum Beispiel, dass Frauen nicht mit Technik umgehen können. Dieser Stereotyp könnte dann ein Grund sein, Frauen als Gruppe abzuwerten im Kontext von Stellenbesetzungen bei einem Technologiekonzern wie zum Beispiel Amazon. Durch solche Stereotypen werden Ungleichheiten begründet und etabliert. Wenn solche Ungleichheiten Grundlage für Entscheidungen sind, können die Effekte dieser Entscheidungen als Diskriminierung – in dem negativ konnotierten Sinne – bezeichnet werden.[136] Diskriminierung bedeutet somit laut Informationswissenschaftlerin Batya Friedman und Technologiephilosophin Helen Nissenbaum eine Schlechterbehandlung einer Person oder Gruppe aus unvernünftigen oder unangemessenen Gründen.[137] Denn wenn es um die Frage geht, ob eine Person mit Technik umgehen kann und dementsprechend geeignet ist für einen Job bei Amazon, ist das Geschlecht irrelevant – somit wäre das Geschlecht der Bewerber:innen ein unvernünftiger Grund für ihre Ablehnung. Allein Faktoren wie das Interesse an Technik, die Ausbildung in dem Bereich oder Arbeitserfahrung sollten relevant und deshalb ein Grund sein, zwischen den Bewerber:innen zu unterscheiden.

133 Krainer 2024:304
134 Zick 2020:8
135 Krainer 2024:304
136 Mittelstadt et al. 2016:8
137 Friedman, Nissenbaum 1996:332

Die unhinterfragte Akzeptanz von Wissen, das aus Datensätzen gewonnen wurde, provoziert deshalb die Möglichkeit der Diskriminierung, weil bei der Analyse von Daten immer Gruppen voneinander unterschieden werden – dies ist die Grundlage von algorithmischer Mustererkennung. Die analysierten Datensätze bilden dabei nicht nur verschiedene Gruppen ab, sondern auch ihre Bewertung. Wenn in der Vergangenheit die Gruppe „Frauen" abgewertet und deshalb nicht eingestellt wurde, dann zeigt sich das in den Daten über bisher erfolgreiche Bewerber:innen. Diese Daten machen eine deskriptive Aussage darüber, welche Kandidat:innen (bzw. meistens Kandidaten) in der Vergangenheit im Bewerbungsprozess erfolgreich waren. Werden diese Daten aber als Grundlage für zukünftige Entscheidungen über die besten Bewerber:innen verwendet, gehen auch alle bisherigen Abwertungen von Gruppen als Status quo ein – so werden Diskriminierungen reproduziert.

Auf diese Weise bilden Datensätze Strukturen ab, die in der Gesellschaft vorherrschen und diskriminierend sein können. Auch wenn ein vermeintlich objektiver Algorithmus die Daten analysiert, findet er in repräsentativen Daten die Muster, die auch in der Gesellschaft zu finden sind. Dabei kann es sich um sexistische Strukturen handeln, weil zum Beispiel bei Amazon weniger Frauen als Männer arbeiten. Bekommt der Algorithmus diese Daten gefüttert, werden seine Ergebnisse dieses Muster repräsentieren. So könnte ein Algorithmus, der dabei helfen soll, die besten Bewerber:innen zu finden, Frauen systematisch ausschließen.[138] Ebenso machen sich rassistische Strukturen bemerkbar, wenn ein Algorithmus bei der Bilderkennung helle Gesichter besser erkennen kann als dunkle Gesichter.[139]

Im Falle der fehlenden Akzeptanz von weiblichen Bewerberinnen bei Amazon reproduziert sich eine gesellschaftliche Ungleichheit und zugeschriebene Ungleichwertigkeit, weil die Daten die Wirklichkeit abbilden. Im Gegensatz dazu entsteht bei der Gesichtserkennung eine Ungerechtigkeit dadurch, dass die Trainingsdaten unvollständig waren und die Gesellschaft nicht vollständig abgebildet haben. Bei der Verarbeitung von Daten gibt es also zwei Möglichkeiten

138 Stahl et al. 2023:10f.
139 Buolamwini, Gebru 2018

der Verzerrungen (Bias).[140] Es besteht zum einen die Gefahr, dass die Daten nicht die Wirklichkeit abbilden und das Ergebnis somit keine Aussage über die Wirklichkeit machen kann. Zum anderen besteht die Gefahr, dass die Daten eine ungerechte Wirklichkeit abbilden, die sich durch ihre Analyse in den Ergebnissen reproduziert.

Wenn man in diesem Kontext die ethische Frage stellt, ob bestimmte Daten zu einem bestimmten Zweck verarbeitet werden sollten, muss zwischen Nutzen und Risiken abgewogen werden. Wenn beispielsweise die Gesundheitsdaten von Personen mit einer chronischen Krankheit analysiert werden, kann womöglich Wissen über die Entstehung der Krankheit oder über die bestmögliche Behandlung extrahiert werden. Denn vielleicht hat sich die Mehrheit der erkrankten Menschen zuvor ungesund ernährt, oder ein Medikament schlägt bei jungen Menschen besser an als bei älteren, aber in Kombination mit einer bestimmten Diät ist es auch bei älteren Menschen wirksam. So kann Prävention gegen die Erkrankung betrieben werden, und es kann die beste Behandlung für alle Erkrankten gefunden werden – die Vorteile der Datenverarbeitung sind hier deutlich zu erkennen. Für die bereits Erkrankten, deren Daten analysiert werden, besteht jedoch durch die Datenverarbeitung das Risiko, dass Unbefugte Zugriff auf ihre Daten bekommen. Denn die Information, dass man eine chronische Krankheit hat, kann zu Schlechterstellung führen, wenn zum Beispiel ein Arbeitgeber die geplante Festanstellung der erkrankten Person überdenkt. Die Risiken, die im Datenökosystem entstehen, werden in Kapitel 3.3 identifiziert und in den Kapiteln 6 und 7 systematisch betrachtet.

Bis hierhin ist erstmal klar, dass das Sammeln und Verarbeiten von Daten sowohl Vorteile hat, als auch Risiken bereithält. Da in dieser Arbeit erörtert werden soll, wie Datenökosysteme ethisch bewertet und dementsprechend gestaltet werden sollten, wird nun definiert, was ein Datenökosystem ist.

140 Eine ausführliche Definition von „Bias“ folgt in Kapitel 7 mit drei verschiedenen Formen der Verzerrung nach Friedman und Nissenbaum (1996).

2.2 Definitionen von Datenökosystemen

Ein Datenökosystem lässt sich als System definieren, das sich zum Ziel setzt Daten zu sammeln und auszuwerten. Der Logiker und Sprachphilosoph Stewart Shapiro definiert jegliches System als „(...) a collection of objects with certain relations."[141] Bei einem Datenökosystem handelt sich also um eine Sammlung von Objekten zwischen denen Daten fließen. In diesem Teilkapitel wird gezeigt, dass die Beziehungen zwischen den Objekten den Datenfluss bestimmen.

Bevor die Idee eines Datenökosystems entstand, war die Nutzung von Daten eine Einbahnstraße, da es keine Rückkopplungsschleife gab zwischen denen, die die Daten bereitgestellt haben und denen, die sie genutzt haben.[142] Dieses Phänomen der Rückkopplung, das ein Datenökosystem von anderweitiger Datenverarbeitung abhebt, wird ausgedrückt durch den Begriff „Ökosystem". Als Mitarbeitende des Bundesministeriums für Wirtschaft und Energie schrieben Thomas Bendig und seine Kolleg:innen der Begriff Ökosystem „(...) greift die Metapher natürlicher Ökosysteme als definierten Lebensraum vielfältiger Organismen und ihrer Umwelt auf."[143] Laut Antti Poikola, der sich als Vorreiter bei der Förderung offen zugänglicher öffentlicher Daten in Finnland bezeichnet, und den Forschern und Designern Petri Kola und Kari A. Hintikka bezieht sich der Begriff des Ökosystems auf ein funktionierendes Ganzes in einem bestimmten Bereich.[144] Konkret funktioniert dieses Ganze als eine zyklische, biologische Umgebung. Laut dem Betriebswirtschaftler Maximilian Heimstädt und seinen Kollegen Fredric Saunderson und Tom Heath, die im Datenschutz Bereich arbeiten, bilden im digitalen Ökosystem die Datensätze sowie die Systeme und Akteure, die diese Daten unterstützen, eine Analogie dazu.[145]

Die Rückkopplung kann sich im Datenökosystem zeigen, indem Datenquellen jedes Mal auch von der Verarbeitung ihrer Daten profitieren, sodass immer ein einmaliger Kreislauf des Datenflusses

141 Shapiro 1997:73
142 Oliveira et al. 2019:590; Pollock 2011
143 Bendig et al. 2019:5; Stand 2025 ist Thomas Bendig Mitarbeiter beim Bundesministerium für Digitales und Staatsmodernisierung.
144 Poikola et al. 2011:13
145 Heimstädt et al. 2014:4

entsteht. Wenn allerdings nicht nur der Zyklus des Prozesses der Datenverarbeitung gemeint ist, sondern auch die zyklische Verwendung von Daten als Ressourcen gefordert wird, stößt die Metapher des Ökosystems bei der Verwendung personenbezogener Daten an rechtliche Grenzen. Denn laut Datenschutzgrundverordnung müssen die Datenquellen, über welche die Daten eine Aussage machen, im Rahmen der Einverständniserklärung zur Verarbeitung ihrer Daten über die Zwecke der Verarbeitung informiert werden.[146] Das heißt, personenbezogene Daten dürfen nicht unangekündigt für andere Zwecke wiederverwendet werden, was den zirkulären Ressourcenfluss erschwert – er müsste vor dem Sammeln der Daten detailliert geplant werden.

Schon bei der Begriffserklärung wird der erste Konflikt deutlich, denn die Idee, dass bereits verwendete Daten in bereinigter, integrierter und verpackter Form zur Wiederverwendung in das Datenökosystem eingebracht werden sollen[147], steht bei der Verwendung von personenbezogenen Daten den Interessen der Personen entgegen. Der eigentliche Sinn eines Datenökosystems ist es aber, dass alle Beteiligten von der Arbeit mit den Daten profitieren.[148] So wird deutlich, dass zwischen den Akteuren Kompromisse erforderlich sind. Das passt aber tatsächlich in den natürlichen Verlauf, da Datenökosysteme sowieso erst durch eine fruchtbare Interaktion zwischen kooperierenden und konkurrierenden Akteuren entstehen.[149] Laut dem Wirtschaftswissenschaftler Ron Adner kann ein Ökosystem entweder als Gemeinschaft von Akteuren oder als Struktur von Aktivitäten, die sich durch ihr Wertversprechen definieren, verstanden werden.[150] In beiden Fällen ist eine koordinierte Zusammenarbeit notwendig, um gemeinsam einen Mehrwert aus den Daten zu ziehen. Das Ziel dieser Zusammenarbeit ist laut Poikola et al. die Verteilung und Entwicklung des Rohstoffes, also der Daten.[151] Ein Datenökosystem besteht demnach aus mehreren vernetzten Akteuren, die Daten wie einen Rohstoff betrachten, der sich idealerwei-

146 Artikel 5, Absatz 1 (b) DSGVO
147 Pollock 2011
148 Poikola et al. 2011:7
149 Heimstädt et al. 2014:1
150 Adner 2017:40
151 Poikola et al. 2011:13

se in einem Kreislauf befindet, dadurch das Ökosystem speist, zur Wertschöpfung beiträgt und allen Beteiligten Vorteile bringt.[152]

Laut Adner spielen insbesondere die Akteure im Datenökosystem eine große Rolle: „The ecosystem is defined by the alignment structure of the multilateral set of partners that need to interact in order for a focal value proposition to materialize."[153] Die Ausrichtungsstruktur bezeichnet das Ausmaß, in dem sich die Akteure über ihre Positionen im System einig sind. Erfolgreich sei ein Ökosystem dann, wenn alle Akteure mit ihren Positionen zufrieden seien. Dabei handele es sich um eine Vielzahl von Akteuren mit multilateralen Beziehungen, die sich nicht auf bilaterale Interaktionen herunterbrechen ließen.[154] Die am System beteiligten Akteure haben die gemeinsame Wertschöpfung zum Ziel. Die Akteure, von deren Beteiligung die geplante Wertschöpfung abhängt, bezeichnet Adner als Partner. Um die Wertschöpfung verwirklichen zu können, werden die dafür erforderlichen Aktivitäten durch das Einbeziehen und Koordinieren von Partnern durchgeführt. Laut Adner ist die gemeinsame Wertschöpfung dabei die Grundlage des Ökosystems.[155]

Adners Definition liegt eine aktivitätsbasierte Perspektive auf die Interdependenzen im Ökosystem zugrunde, die er die Ökosystem-als-Struktur Perspektive nennt[156]: „(…) the ecosystem-as-structure view begins [with] the value proposition, considers the activities required for its materialization, and ends with actors that need to be aligned (…)."[157] Aus diesem Blickwinkel beginnt ein Ökosystem also mit der Absicht einer Wertschöpfung und strebt das Identifizieren der Gruppe von Akteuren an, die interagieren müssen, um das Vorhaben zu verwirklichen.[158]

Eine andere mögliche Perspektive wäre laut Adner der akteurszentrierte Ökosystem-als-Zugehörigkeit Ansatz: „The ecosystem-as-affiliation approach begins with the actors (usually defined by their ties to a focal actor), considers the links among them, and ends with the possible value propositions and enhancements that the ecosys-

152 Zubcoff et al. 2016:251f.
153 Adner 2017:40, 42
154 Adner 2017:42
155 Ebd.:43
156 Ebd.:41
157 Ebd.:44
158 Ebd.:41

tem can generate."[159] Hierbei stehen die Akteure eines Ökosystems im Mittelpunkt, aber es werden nicht ihre Aktivitäten betrachtet. Dieser Ansatz liegt zum Beispiel der Definition eines staatlichen Ökosystems von der Kommunikationswissenschaftlerin Teresa M. Harrison und ihren Kolleginnen Theresa A. Pardo und Meghan Cook aus dem Center for Technology in Government zugrunde: „Our perspective on the *open government ecosystem* envisions government organizations as central actors, taking the initiative within networked systems organized to achieve specific goals related to innovation and good government."[160] Allerdings dürfen die Aktivitäten der Akteure meiner Meinung nach in der Definition nicht fehlen, weil Akteure sich durch ihre Aktivitäten auszeichnen und ohne diese auch keine Wertschöpfung betreiben können.

Unabhängig davon, ob Ökosysteme aus der Struktur- oder Zugehörigkeits-Perspektive betrachtet werden, spielen die Akteure eine wichtige Rolle, da die Wertschöpfung im Datenökosystem in einem System von Akteuren stattfindet.[161] Poikola et al. stimmen zu: „The ecosystem sees organisations and private people as actors, interacting with each other to mould the conventions of [the] ecosystem."[162] Die Elemente eines Datenökosystems sind Ressourcen, Aktivitäten, Akteure, Rollen, Positionen im Fluss der Aktivitäten und Beziehungen zwischen den Akteuren.[163] Die Informatiker:innen Marcelo Iury S. Oliveira, Glória de Fátima Barros Lima und Bernadette Farias Lóscio beschreiben die Akteure eines Datenökosystems folgendermaßen: „Actors are autonomous entities such as enterprises, institutions, or individuals, which act in the Data Ecosystem in one or more specific roles."[164] Die Akteure werden durch eine Reihe von Interessen motiviert und haben in der Regel eine Verpflichtung gegenüber dem Datenökosystem – dadurch können sie einen Anreiz haben für ihre Aktivitäten.[165] Jeder Akteur hat eine Rolle in den Informations-, Material- und Geldflüssen und diese beeinflussen die

159 Adner 2017:43f.
160 Harrison et al. 2012:907
161 Immonen et al. 2014:88
162 Poikola et al. 2011:28
163 Adner 2017:43f.; Oliveira, Lóscio 2018:8
164 Oliveira et al. 2019:606
165 Ebd.

Beziehungen untereinander.[166] Oliveira et al. definieren den Begriff der Rolle so: „A role is a function played by an actor in the ecosystem. It is related to a position and a set of duties."[167] Gewisse Pflichten entstehen aus der spezifischen Rolle und Position des Akteurs im Datenökosystem.

Ein Datenökosystem ist also ein soziotechnisches System, das gemeinsam die Verarbeitung von Daten anstrebt, um einen Mehrwert aus den Daten zu ziehen.[168] Dieses soziotechnische System besteht aus Menschen und Organisationen einerseits, Infrastruktur und Technologie andererseits, und seine Inhalte und Vorgehensweisen werden durch Ressourcen und Richtlinien bestimmt. Die wichtigsten Ressourcen in einem Datenökosystem sind dabei Datensätze.[169] Zu der Verarbeitung von Daten gehören Sammlung, Integration, Analyse, Speicherung, gemeinsame Nutzung, Datensicherheit und -schutz sowie Archivierung. Ein Mehrwert wird aus den Daten gezogen, wenn durch ihre Verarbeitung zum Beispiel Dienstleistungen besser erbracht werden können und wenn dadurch eine Politik geschaffen wird, die Bürgern, Unternehmen und staatlichen Stellen zugutekommt.[170] Diese erforschende, wertschöpfende Verarbeitung von Daten ist nur möglich durch die Zusammenarbeit der verschiedenen Akteure im Datenökosystem. Diese Akteure – der soziale Teil des soziotechnischen Systems – sind für das Produzieren, Bereitstellen oder Konsumieren der Daten zuständig, und sie sind dabei durch Beziehungen miteinander verbunden.[171] In der Literatur finden sich unterschiedliche Definitionen von Datenökosystemen, aber die Konzepte der Akteure, Rollen, Beziehungen und Ressourcen sind allen gemeinsam.[172]

Ein weiterer interessanter Begriff ist der des Open Data Ecosystems. Der Begriff „Open Data" bezieht sich auf die offene Zugänglichkeit der Daten innerhalb eines Datenökosystems und wird von der Open Knowledge Foundation, eine Stiftung, die sich für digital allen Menschen zugängliches Wissen einsetzt, folgendermaßen defi-

166 Immonen et al. 2014:88
167 Oliveira et al. 2019:606
168 Shah et al. 2020:112; Oliveira et al. 2019:590
169 Oliveira, Luscio 2018:7
170 Shah et al. 2020:112
171 Oliveira et al. 2019:604
172 Ebd.:604

niert: „Open data and content can be freely used, modified, and shared by anyone for any purpose".[173] Die Offenheit der Daten ist dabei nicht als ihre inhaltliche Transparenz zu verstehen, sondern allein als unbegrenzter Zugang zu den Daten.[174] Bei Open Data Ecosystems besteht eine der Leistungen darin, die Daten öffentlich bereitzustellen. Hier passt die Beschreibung von sogenannten „Data spaces" (Datenräume) aus dem Data Governance Act, eine Verordnung der Europäischen Union zur Erleichterung der gemeinsamen Nutzung und Wiederverwendung öffentlich zugänglicher, geschützter Daten über Sektoren und EU-Länder hinweg. Diese Datenräume sind definiert als Datenökosysteme in einem bestimmten Anwendungsbereich, die auf gemeinsamen Regeln basieren. Solche Datenräume enthalten Daten und ermöglichen allen Nutzenden den Zugriff darauf. Dabei werden von einem Akteur mehrere Datenökosysteme gleichzeitig genutzt, sodass sie sich in und zwischen verschiedenen Datenräumen bewegen.[175]

In der Idealvorstellung von Datenräumen haben die Akteure die Kontrolle über ihre eigenen Daten. So schreiben der CEO von der International Data Spaces Association, Lars Nagel, und der Senior Vizepräsident von Innopay, Douwe Lycklama: „Data will only be shared and exchanged between actors if they decide to do so."[176] Daten können zum Beispiel geteilt werden, weil sich die Akteure davon einen finanziellen Gewinn versprechen. Aber es muss dabei nicht um eine finanzielle Transaktion gehen.[177] Es kann auch mit Daten gehandelt werden, wenn diese für einen Akteur im Ökosystem einen Wert haben oder um die Qualität eines Produktes bzw. einer Dienstleistung zu verbessern.

An einigen Stellen wird Open Data gepriesen, weil sich der offene Zugang zu Daten positiv auf die Wertschöpfung im Datenökosystem auswirkt, weil er Bürgerbeteiligung sowie Innovation fördert, weil er das Wirtschaftswachstum ankurbelt, weil er die Bereitstellung öffentlicher Dienstleistungen verbessert und dabei Kosten spart.[178] Die

173 Open Knowledge Foundation 2022; Hervorhebungen entfernt
174 Yu, Robinson 2012:208
175 Nagel, Lycklama 2021:7
176 Ebd.:92
177 Steffen et al. 2021:46
178 Zuiderwijk et al. 2016:223; Zuiderwijk et al. 2014:17; Davies 2011:1; Zeleti, Ojo 2014:477

Offenheit von Daten im Allgemeinen wird sogar gefordert, zum Beispiel von dem Informatiker Mark A. Parsons und seinen Kollegen: „Data should generally be openly accessible."[179] Parsons et al. versprechen sich davon, dass die Interaktion zwischen den Akteuren im Datenökosystem dadurch gefördert wird und dass infolgedessen die Zusammenarbeit und das interdisziplinäre Verständnis verbessert werden.[180] Manche Autor:innen argumentieren, dass der öffentliche Zugang zu Daten notwendig ist, um die Transparenz der Regierung zu ermöglichen[181] und dass Open Data dadurch auch eine stärkere demokratische Rechenschaftspflicht möglich macht[182]. Hierbei würde es sich dann um ein sogenanntes Open Government handeln, das durch frei zugängliche, wiederverwendbare und weiterverbreitete Daten seinen eigenen Zielen gerecht werden kann.[183]

Allerdings kommt der Sozialwissenschaftler Tim Davies zu dem Schluss, dass Open Data vor allem gut organisierte Interaktion zwischen den Akteuren im Datenökosystem fordert. Gute Zusammenarbeit im Rahmen einer stabilen Infrastruktur entsteht nicht von alleine, sondern müsse mit Hilfe von Koordinierungsstrategien erarbeitet werden.[184] Außerdem macht der Begriff Open Data an sich keine Aussage über den Inhalt der Daten, sondern nur über ihren technischen und rechtlichen Zustand.[185] Deshalb kann allein aus dieser technischen Beschreibung der Zugangsmöglichkeiten zu Daten noch kein Schluss gezogen werden über den tatsächlichen oder erwarteten Nutzen der Datenweitergabe.[186]

Datenökosysteme entstehen auch in der deutschen öffentlichen Verwaltung. Denn das Onlinezugangsgesetz (OZG) verpflichtet die deutsche Verwaltung dazu, ihre Services auch in digitaler Form anzubieten. Laut einem Bericht des Bundesministeriums für Wirtschaft und Energie (BMWi) über das Projekt Gaia-X setzt ein solcher digitaler Wandel „(...) eine leistungsfähige Cloud-Infrastruktur voraus

179 Parsons et al. 2011:557
180 Ebd.:565
181 Zuiderwijk et al. 2014:17; Geiger, Lucke 2012:265; Zubcoff et al. 2016:250
182 Davies 2011:1; Zeleti, Ojo 2014:477
183 Zubcoff et al. 2016:250
184 Davies 2011:5
185 Heimstädt et al. 2014:2; Yu, Robinson 2012:188f., 206
186 Yu, Robinson 2012:208

(...)."[187] Durch das europäische Projekt Gaia-X soll eine vernetzte Dateninfrastruktur geschaffen werden. Zu diesem Zweck werden verschiedene Projekte durch das Bundesministerium für Wirtschaft und Klimaschutz gefördert – darunter auch das in der Einleitung erwähnte MERLOT Projekt.[188]

Ziel dieser Datenökosysteme ist es, das Teilen von Informationen so leicht zu machen, dass die Dienstleistungen der Verwaltung schneller erbracht werden können und mit weniger Aufwand für die Antragstellenden verbunden sind. Bei der Beantragung von Elterngeld muss beispielsweise der Einkommensnachweis der Eltern eingereicht werden.[189] Damit die Eltern nicht ihre Einkommensnachweise suchen und damit zu einem Rathaus gehen müssen, sollen die Einkommensnachweise aus anderen Kontexten genutzt werden können. Wenn die Eltern vorher schon Bürgergeld oder Bafög beantragt haben, mussten sie dafür ebenfalls ihr Einkommen angeben. Das Datenökosystem soll ermöglichen, dass die antragstellenden Eltern der Elterngeldbehörde den Zugriff auf ihre Einkommensnachweise aus dem Bürgergeld- oder Bafög-Antrag ermöglichen. Dadurch würden die Daten der Antragstellenden (Datenquellen) zu verschiedenen Zwecken verarbeitet, und sowohl die Antragstellenden selbst als auch die bearbeitenden Behörden profitieren davon durch einen geringeren Aufwand und Zeitersparnis.

Der Begriff „open data" könnte sich im Kontext der öffentlichen Verwaltung zum einen darauf beziehen, dass die Daten des Ökosystems allgemein zugänglich sind oder zum anderen auf die Bereitstellung der Daten für den öffentlichen Sektor, sodass die Daten verwaltungsintern zugänglich sind. Die Daten, die allgemein zugänglich gemacht werden könnten, wären Daten aus der Verwaltung oder aus den verschiedenen Ministerien und Ämtern. Auf diese Weise könnten Daten aus verschiedenen Bereichen miteinander verknüpft werden. So könnten zum Beispiel Daten über das Einkommen einer Person aus dem Bürgergeldantrag beim Bundesarbeitsministerium weitergeleitet werden an das Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen für den Wohngeldantrag der gleichen Person.

187 Bundesministerium für Wirtschaft und Energie 2019:34

188 Mehr Informationen unter: https://www.bmwk.de/Redaktion/DE/Dossier/gaia-x.html

189 Bundesministerium der Finanzen 2023

Daraufhin stellt sich die Frage, ob auf Grund des erwarteten Nutzens der Datenweitergabe in Datenökosystemen alle Daten als Open Data öffentlich zugänglich sein sollten. Die erwarteten positiven Auswirkungen von offen zugänglichen Daten reichen von einer Steigerung der Wertschöpfung im Datenökosystem über eine Verbesserung von Dienstleistungen und Förderung von Bildung bis hin zur Schaffung neuer Innovationen und Erreichung von mehr Transparenz in der Demokratie.[190] Sobald es sich aber um personenbezogene Daten handelt, wird es eine Abwägung zwischen den Vorteilen und dem Schutz der Privatsphäre der Datenquellen geben müssen.

Abschließend kann man sagen, dass sich ein Datenökosystem über den Datenfluss zwischen Akteuren mit verschiedenen Rollen definiert. Zudem wurde wieder deutlich, dass insbesondere der offene Umgang mit digitalen Daten viele Vorteile verspricht. Es wurde aber auch angedeutet, dass dieser Umgang mit den Daten für bestimmte Akteure auch Risiken bereithält.

2.3 Akteure im Datenökosystem

Um ein besseres Verständnis von den Akteuren im Datenökosystem zu bekommen, stelle ich ein eigenes Modell von der Struktur eines Datenökosystems auf. Die Struktur definiert sich dabei laut Shapiro wie folgt: „A *structure* is the abstract form of a system, highlighting the interrelationships among the objects, and ignoring any features of them that do not affect how they relate to other objects in the system."[191] Das heißt, das Modell stellt nur die abstrakte Form eines Datenökosystems dar, indem die Beziehungen zwischen den Objekten hervorgehoben werden. Die Objekte sind hier die menschlichen Akteure, da die technologischen Objekte des Datenökosystems selbst keine Beziehungen eingehen können – sie werden nur benutzt durch die menschlichen Akteure. In diesem Modell der Struktur eines Datenökosystems werden also nur die Merkmale der Objekte dargestellt, die einen Einfluss darauf haben, wie sie sich zu anderen Objekten im System verhalten.

190 Yu, Robinson 2012:182; Poikola et al. 2011:11, 14; Shah et al. 2020:112
191 Shapiro 1997:74

Das Modell von der Struktur eines Datenökosystems soll auf Grundlage der unter 2.2 erfolgten Definition folgende Kriterien erfüllen:

1) Der Datenfluss soll klar erkennbar sein.
2) Die beteiligten Akteure sollen voneinander differenziert sein.
3) Die Rollen der Akteure sollen klar ersichtlich sein.

Es gibt bereits Modelle von Datenökosystemen, doch sie alle erscheinen mir unübersichtlich oder sogar unvollständig. Entweder fehlt die Bezeichnung der Akteure, sodass unklar ist, wer die Aufgaben im Datenökosystem ausführt[192], oder essentielle Schritte – wie die Datenverarbeitung – werden nicht separat betrachtet, sondern zum Beispiel den Nutzenden als Aufgabe zugeschrieben[193]. In einem anderen Modell ist es jedem Menschen möglich einen eigenen Datensatz oder Algorithmus zur Verfügung zu stellen[194], was das Überprüfen der Qualität erschweren würde.

Suwook Ha, Seungyun Lee und Kangchan Lee, die am Electronics and Telecommunications Research Institute forschen, haben 2014 ein sehr übersichtliches Modell erstellt mit Datenprovidern, Serviceprovidern und Datenkonsumierenden, welches in Abbildung 1 zu sehen ist.

Die Aufgabe der Datenprovider ist es, neue Datensammlungen in das Datenökosystem einzuführen, damit diese verarbeitet werden können. Der Serviceprovider bietet die dafür notwendigen Hilfsmittel an: Infrastrukturen für den Verbraucher und Analyseanwendungen. Die Analyseanwendungen ermöglichen Datenvisualisierung und -analyse, und die Infrastruktur ermöglicht die Sammlung, Speicherung und Vorverarbeitung der Daten. Die Datenkonsumierenden nutzen die Ergebnisse der Datenanalyse und können diese außerhalb vom Datenökosystem zur Verfügung stellen.[195] Dieses Modell zeigt den Verlauf der Daten innerhalb des Verarbeitungsprozesses klar auf (Kriterium 1), es wird jedoch nicht erkennbar, woher die Daten kommen. Das Fehlen der Datenquellen als Akteur halte ich für ein Problem, weil dadurch der Schutz der Datenquellen bei den Abläufen in einem Datenökosystem keine Rolle zu spielen scheint.

192 Davies 2011:3; Attard et al. 2016:454f.
193 Dawes et al. 2016:39
194 Traub et al. 2020:8
195 Ha et al. 2014:166

Es ist so nicht möglich, die Position der Datenquellen zu problematisieren.

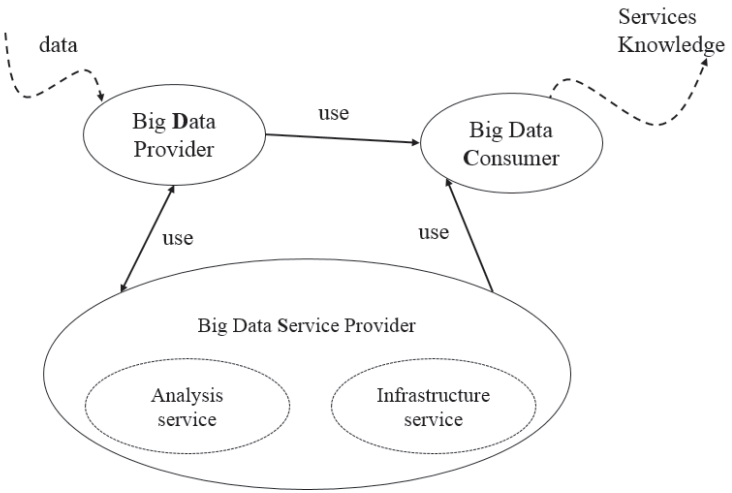

Abbildung 1 Big Data Ökosystem von Ha et al.[196]

Das gleiche Defizit haben die bereits erwähnten Technologieforscher Syed Iftikhar Hussain Shah, Vasilis Peristeras und Ioannis Magnisalis in ihrem Modell aus dem Jahr 2020. Sie beschreiben vier „Elemente" eines Datenökosystems und identifizieren ihre jeweilige Rolle und Motivation.[197] Ich würde sagen Shah et al. beschreiben nicht die Elemente eines Datenökosystems (Ressourcen, Aktivitäten, Akteure, Rollen, Positionen im Fluss der Aktivitäten und Beziehungen zwischen den Akteuren[198]), sondern sie beschreiben vier Akteure eines Datenökosystems. Demnach sind die vier Akteure eines Datenökosystems *Data Publishers*, die Daten kostenlos bereitstellen oder verkaufen, eine *Data Business Entity*, welche sowohl eigene Daten beisteuert und diese analysiert als auch die beste Datenquelle für die Bedürfnisse der Data User findet, *Data Support Serviceprovider*, die

196 Ha et al. 2014:166
197 Shah et al. 2020:106ff.
198 Adner 2017:43f.; Oliveira, Lóscio 2018:8

gegen Geld einen Speicherort für die Daten bereitstellen, und *Data Users*, die dank der Analyse der Daten Wert daraus schöpfen.[199] Eine der Aufgaben der Data Business Entity ist es zwar, bestmögliche Datenquellen zu finden, dabei sind aber die Datenquellen kein Akteur des Datenökosystems. Aus meiner Sicht erfüllen die Modelle von Ha et al. und Shah et al. dadurch weder Kriterium 2 noch 3.

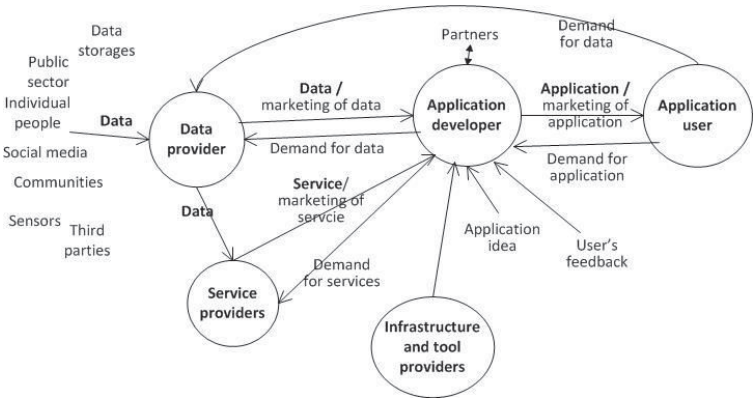

Abbildung 2 Offenes Datenökosystem von Immonen et al.[200]

Anders ist das bei Anne Immonen, Marko Palviainen und Eila Ovaska aus dem VTT Technical Research Centre of Finland – sie schließen in ihrem Modell Datenquellen ein, wie in Abbildung 2 erkennbar ist. Die Daten gelangen zum Beispiel von Individuen oder Sozialen Medien zu den Datenprovidern. Diese stellen die Daten wiederum Serviceprovidern zur Verfügung, sowie denjenigen, die die Anwendungen entwickeln (Application developer). Die Serviceprovider produzieren daraufhin datenbezogene Dienste. Mit Hilfe dieser Dienste und angebotener Infrastruktur entwickeln die Application developer Anwendungen für die Daten. Diese werden von den Anwendungsnutzenden (Application user) konsumiert. Das heißt, die Anwendungen nutzen Daten und Dienste und produzieren wertvolle Informationen und Wissen für die Nutzenden.[201] Obwohl dieses Modell in seiner Vollständigkeit überzeugt, erscheint

199 Shah et al. 2020:106–109
200 Immonen et al. 2014:91
201 Ebd.:91

mir der Datenfluss über die Dienstanbietenden unnötig kompliziert. Da Dienst und Anwendung nicht separat entwickelt werden müssen, ist es problemlos möglich, den Datenfluss im Modell zu vereinfachen. Das Modell von Immonen et al. erfüllt also Kriterium 2 und 3, da die Akteure und ihre Rollen klar erkennbar sind. Kriterium 1 wird jedoch nicht erfüllt, weil der Datenfluss sehr unübersichtlich abgebildet ist.

In der Literatur kommen viele interessante Ideen zusammen, sie vermischen sich allerdings auch – hier muss begriffliche und definitorische Klarheit geschaffen werden. Typische Rollen, die in den meisten Modellen eines Datenökosystems auftauchen, sind Datenproduzierende, Datenanbietende und Datennutzende.[202] An dieser Stelle sind die Begrifflichkeiten schon missverständlich. Sind Datenproduzierende die Quellen der Daten oder diejenigen, die Daten zusammentragen und in diesem Sinne Datensätze produzieren? Bieten Datenanbietende die Daten zur Analyse an oder zur Nutzung nach der Verarbeitung? Aus meiner Sicht wird es klarer, wenn man von den Akteuren *Datenquellen*, *Datengenerierenden*, *Datenverarbeitenden* und *Datennutzenden* spricht. Dabei handelt es sich um Personen, die ihre Daten zur Verfügung stellen (*Datenquellen*), verschiedene Beteiligte, die die Daten sammeln (*Datengenerierende*), diejenigen, die die Daten kombinieren und analysieren, um sie dann aufbereitet zur Verfügung stellen zu können (*Datenverarbeitende*) und die Personen, die mit den aufbereiteten Daten arbeiten (*Datennutzende*). Diese Rollen beschreiben zugleich die vier wichtigsten Akteure eines Datenökosystems, weil sie alle direkt am Datenfluss beteiligt sind.

Welche Aktivitäten von welchen Akteuren ausgeführt werden, ist in der Literatur nicht eindeutig und wird deshalb an dieser Stelle erörtert. Laut einigen Modellen gibt es einen Keystone Actor – einen Akteur, der die Schlüsselrolle spielt, weil er für Stabilität sorgt und eine treibende Kraft im Datenökosystem darstellt.[203] In vielen Modellen wird er durch die Regierung repräsentiert, dabei handelt es sich dann um sogenannte Open Government Data Ecosystems.[204] Keystone Actors sind laut den Informatikerinnen Bonnie A. Nardi

202 Oliveira et al. 2019:606
203 Ebd.:608
204 Oliveira, Loscio 2018:7

und Vicki L. O'Day „(...) skilled people whose presence is necessary to support the effective use of technology."[205] Diese Menschen fungieren als Übersetzende, Moderierende, Lehrende und Vermittelnde, indem sie verschiedene Institutionen und unterschiedliche Disziplinen zusammenbringen.[206] Der keystone actor verwaltet also das Datenökosystem und sollte laut Oliveira et al. deshalb formale Richtlinien, Regeln und Praktiken fördern sowie Verantwortung für Überwachung, Bewertung, Entscheidungsfindung und das Ergreifen von Maßnahmen übernehmen.[207] Nach der Ansicht der Technologieforscher:innen François van Schalkwyk, Michelle Willmers und Maurice McNaughton ermöglicht der keystone actor den Datenfluss durch seine Vermittlung zwischen Institutionen und Disziplinen – aber er ist keine notwendige, treibende Kraft: „Keystone species are enablers, not necessarily drivers in the ecosystem; they can be useful but they are not essential to the sustained functioning of an ecosystem."[208] Ich nenne diejenigen, die im Datenökosystem zwischen Institutionen und Disziplinen vermitteln sowie Verantwortung tragen für Regeln, Überwachung und Maßnahmen, im Folgenden die Verwaltenden. Dieser fällt keine bedeutende ethische Funktion im Datenökosystem zu, aber sie ist notwendig, um ein Datenökosystem aus einer technischen Perspektive vollständig darzustellen.

Laut dem Betriebswirtschaftler Marco Iansiti und dem Technologieberater Roy Levien sowie Harrison et al. fällt zudem das Erstellen einer technischen Plattform zum Teilen der Daten in den Aufgabenbereich des keystone actors.[209] Aus meiner Sicht überschneidet sich dies mit den Aufgaben der sogenannten Serviceprovider, welche Oliveira et al. so beschreiben: „The service provider is responsible for producing and supporting software resources such as tools, applications, services, libraries, or other software products (...)."[210] Ein Service im Datenökosystem unterstützt die Analyse von Daten, um zur Wertschöpfung beizutragen.[211] Deshalb ist es die Aufgabe der Serviceprovider, zum einen die Infrastruktur für das Datenöko-

205 Nardi, O'Day 1999:53
206 Ebd.
207 Oliveira et al. 2019:608, 609, 610
208 van Schalkwyk et al. 2016:77
209 Iansiti, Levien 2004; Harrison et al. 2012:911
210 Oliveira et al. 2019:608
211 Immonen et al. 2014:89

system bereitzustellen und zum anderen die Analyseergebnisse für die Datennutzenden zu visualisieren. Die Bereitstellung der Infrastruktur ermöglicht die Datenspeicherung sowie das Erstellen und Verwalten der digitalen Plattform, auf der die verarbeiteten Daten für Datennutzende zur Verfügung gestellt werden können.[212] Die Wissenschaftler Juho Lindman, Tomi Kinnari und Matti Rossi aus der angewandten Informatik sind der Auffassung, dass die Aufgaben eines Serviceproviders von Unternehmen übernommen werden und zudem die Beratung von Kunden bezüglich ihrer Möglichkeiten zur Nutzung der Daten umfassen.[213] Für ein erfolgreiches Datenökosystem müssten Serviceprovider zudem die Informationen aktuell und zugänglich halten und dafür technische Services anbieten.[214] Im Folgenden nenne ich diejenigen, die die Analyse von Daten durch technische Services unterstützen, Infrastrukturbereitstellende.

Immonen et al. schreiben, dass Serviceprovider aus den Daten Informationen und Wissen produzieren.[215] Das ist aber nicht ihre Aufgabe, denn ein Serviceprovider unterstützt zwar die Analyse der Daten durch das Verwalten der Infrastruktur, aber er analysiert nicht selbst. Dafür gibt es die separate Rolle der Analysten, deren Funktion in der Literatur oft auch als Service beschrieben wird. Die Informatiker:innen Ying Chen, Jeffrey Kreulen, Murray Campbell und Carl Abrams zum Beispiel führen im Jahr 2011 Analytics as a Service (AaaS) provider als neue Entität im Datenökosystem ein.[216] Wer eine Analyse als Service bereitstellt, bietet technische Werkzeuge sowie die Durchführung der Analyse selbst an.[217] Ein solcher Service kann laut Immonen et al. aus Eingabedaten relevante Daten für einen bestimmten Kontext oder Bereich erzeugen. Diese kontextspezifischen Daten, die durch die Analyse entstehen, sollen bedeutsame Informationen sein – dadurch entstehe die Wertschöpfung im Datenökosystem.[218] Die Bereitstellenden von Analysen als Service bieten zwar im weitesten Sinne einen Service an, indem ihre

212 Ha et al. 2014:166, 168, 169; Lindman et al. 2016:5
213 Lindman et al. 2016:5
214 Ebd.:9
215 Immonen et al. 2014:95
216 Chen et al. 2011:14
217 Ebd.:12, 15, 17
218 Immonen et al. 2014:89

Datenanalyse eine Dienstleistung sein kann, jedoch ist ihre Rolle im Datenökosystem die der Datenverarbeitenden.

Chen et al. haben außerdem den Data as a Service (DaaS) provider als Akteur im Datenökosystem identifiziert. Sie sammeln, generieren und aggregieren Daten, bieten aber auch Datenzugriff, -verarbeitung und -verwaltungsdienste an.[219] Wer Daten als Service anbietet, sammelt und verwaltet viele verschiedene Datenquellen, sodass rohe Daten zur Verarbeitung vorbereitet und den Analysten angeboten werden können.[220] Solange sich die Arbeit mit den Daten auf die Vereinheitlichung des Datensatzes zur Vorbereitung für Datenverarbeitende beschränkt, fällt dies unter die Aufgaben der der Datengenerierenden.

Daten zur Nutzung frei geben, Daten generieren und Daten nutzen können viele Menschen – Daten analysieren können nur wenige. Deshalb ist die Funktionsfähigkeit eines Datenökosystems überaus abhängig von denen, die die Daten analysieren können. Darum bilden die Datenverarbeitenden das Herzstück eines Datenökosystems. Allerdings kann ihr Zugang zum Datenökosystem und auch dessen Sicherheit gefährdet werden, wenn die Infrastruktur von Dienstanbietenden außerhalb der rechtlichen oder territorialen Zuständigkeit der Datenverarbeitenden erstellt und verwaltet wird.[221] Deshalb ist es ratsam, dass die Verantwortung für die Infrastruktur und die Datenverarbeitung in einer Hand liegt, indem sie zum Beispiel von Mitarbeitenden des gleichen Unternehmens übernommen wird.

Die Verwaltung, die Verbindungen zwischen den Akteuren herstellt, könnte man im Datenökosystem da vermuten, wo am meisten von der Vermittlung profitiert wird. Es ist durchaus denkbar, dass die Datenquellen auch die Datennutzenden sein wollen und also ein großes Interesse an der Verarbeitung ihrer Daten haben. Es kann aber auch sein, dass die Analysten aus dem Team der Datenverarbeitenden einen Algorithmus erstellt haben, der viele Daten benötigt, um zu lernen und sich zu entwickeln. In dem Fall liegt den Datenverarbeitenden viel an der Vermittlung zwischen den Akteuren im Datenökosystem, da sie zum Training ihres Algorithmus darauf angewiesen sind, dass die Datengenerierenden ihnen

219 Chen et al. 2011:14
220 Ebd.:12, 14, 17
221 Shin 2016:842

Daten der Datenquellen zuspielen. Die Funktion der Verwaltung als Übersetzende für die effektive Nutzung von Technik[222] ist ebenfalls sehr nützlich für die Datenverarbeitenden: Jemand, der die Daten versteht, muss erklären, welches Potenzial der Datensatz hat und jemand, der weiß, was Datennutzende wollen, muss das Ziel in einen technischen Auftrag umwandeln können. Programmierende müssen verstehen, welche Fragen beantwortet werden sollen, und diejenigen, die verarbeitete Daten für Datennutzende bereitstellen, müssen die Antworten des Algorithmus verstehen. Die Vermittlung zwischen Datenquellen und Datenverarbeitenden übernehmen hingegen die Datengenerierenden. Wenn es also um die Vermittlung von Daten geht, liegt diese Tätigkeit bei den Datengenerierenden.
Wenn man aber betrachtet, wer Verantwortung tragen sollte für Regeln, Überwachung und Maßnahmen im Datenökosystem[223], wird deutlich, dass die Verwaltenden im Idealfall unabhängig von den Akteuren agieren, sodass niemand bevorzugt oder benachteiligt wird. Aber welcher der Akteure sorgt dafür, dass das Datenökosystem verwaltet wird? Ich behaupte, es ist der Akteur, der die Entstehung des Datenökosystems zu einem bestimmten Zweck anstößt.

Da bisher keins der vorgestellten Modelle aus der Literatur alle der von mir aufgestellten Kriterien erfüllt, erstelle ich an dieser Stelle ein neues Modell der Struktur eines Datenökosystems. Das heißt, es werden die Objekte dargestellt, welche das System ausmachen und es wird die Beziehung zwischen den Objekten hervorgehoben.[224] Dabei werden außerdem alle Elemente eines Datenökosystems dargestellt.[225] Denn die Akteure sind die Objekte des Systems. Nicht alle der bisher besprochenen Akteure, wie zum Beispiel die Verwaltung, haben ethische Relevanz und damit Relevanz für diese Arbeit.[226] Deshalb werde ich mich im Folgenden auf die vier Akteure mit ethischer Relevanz konzentrieren: Personen, die ihre Daten zur Verfügung stellen (*Datenquellen*), verschiedene Beteiligte, die die Daten generieren (*Datengenerierende*), diejenigen, die die Daten kombinieren und analysieren, um sie dann aufbereitet zur Verfügung stellen

222 Nardi, O'Day 1999:53
223 Oliveira et al. 2019:608, 609, 610
224 Shapiro 1997:74
225 Adner 2017:43f.; Oliveira, Lóscio 2018:8
226 Was ethische Relevanz hier bedeutet wird im Laufe der Arbeit deutlich.

zu können (*Datenverarbeitende*) und die Personen, die mit den aufbereiteten Daten arbeiten (*Datennutzende*).

Besagte *Datenquellen* sind all die Akteure, die Daten erzeugen. Diese Daten können zum Beispiel durch Messungen erzeugt werden, aber jede Person ist auch an sich eine Datenquelle. Denn Daten, die sich auf die eigene Person beziehen, erzeugen wir Menschen allein durch unsere Existenz. So sind zum Beispiel Name und Wohnort Daten, mit denen eine Person identifiziert werden kann, aber auch das Verhalten im Internet erzeugt Daten, die persönliche Präferenzen abbilden können. Deshalb kann prinzipiell jede Person eine Datenquelle sein. Der Datenfluss im Datenökosystem beginnt aber erst dann, wenn die Daten für *Datengenerierende* von Interesse sind. Denn Datengenerierende sammeln und speichern die Daten, meist schon zum Zweck der Verarbeitung. Entweder sie sind dabei initiativ und suchen selber nach für ihr Projekt passenden Daten, oder sie bieten gemeinsam mit den Datenverarbeitenden gewisse Ergebnisse an, wenn Datenquellen bestimmte Daten über sich preisgeben: Sie sollen zum Beispiel ihre Schulnoten angeben und bekommen eine Berufsempfehlung, oder sie sollen unter anderem Einkommen und Wohnungsgröße angeben und erfahren, ob sie Anspruch auf Wohngeld haben. Damit diese versprochenen Ergebnisse zustande kommen, braucht man die *Datenverarbeitenden*. Sie kombinieren und analysieren die von den Datengenerierenden gesammelten Daten mit Hilfe von Algorithmen und bereiten die algorithmischen Ergebnisse auf, um sie den Datennutzenden zur Verfügung zu stellen. Welche Aussagekraft die Ergebnisse haben, hängt dabei, wie oben erläutert, von den analysierten Daten ab. Die *Datennutzenden* werden mit den Ergebnissen konfrontiert. Wenn sie auch schon bewusst als Datenquelle gedient haben, ist ihnen klar, dass die Ergebnisse eine Aussage über sie machen und ihnen gewissen Möglichkeiten bieten oder verwehren können. Wenn die Daten der Datenquellen aber ohne ihr Zutun gesammelt und verarbeitet wurden, wie zum Beispiel durch Cookies bei der Internetbenutzung, dann ist es auch möglich die Datenquellen durch personalisierte Werbung zu beeinflussen – Datennutzende sind in dem Fall dann Werbende, welche die Ergebnisse nutzen, um ihre Produkte bei den Datenquellen anzupreisen, denen sie gefallen könnten.

Hier wird deutlich, dass Daten die Ressourcen des Datenökosystems darstellen. Außerdem wird durch die Beschreibung erkennbar,

dass jeder Akteur eine gewisse Rolle im System hat, sodass jedem Akteur bestimmte Aktivitäten zugeschrieben werden können. So entstehen im Fluss der Aktivitäten gewissen Positionen im System, da manche Akteure mehr Kontrolle über das Vorgehen im Datenökosystem haben als andere. All diese Elemente bedingen die Beziehungen zwischen den Akteuren.[227]

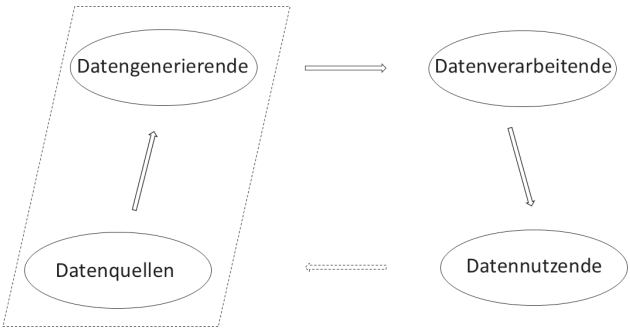

Abbildung 3 Ethisch relevante Akteure eines Datenökosystems[228]

Mein Modell der ethisch relevanten Akteure eines Datenökosystems in Abbildung 3 zeigt, dass die Daten von den Datenquellen zu den Datengenerierenden fließen, von da aus zu den Datenverarbeitenden und nach der Analyse gelangen sie weiter zu den Datennutzenden. Dabei können Datenquellen und Datengenerierende auch ein Akteur sein, was durch den gestrichelten Rahmen in Abbildung 3 dargestellt wird. Wenn zum Beispiel Hersteller von Flugzeugteilen Wartungsdaten erzeugen, sammeln und weitergeben, dann sind diese Hersteller gleichzeitig Datenquellen und Datengenerierende.[229] Dieses Detail ist wichtig, weil nur bei separaten Datenquellen und Datengenerierenden ein ungleiches Kontrollverhältnis zwischen diesen Akteuren bestehen kann und nur dann an dieser Stelle eine ethische Gefahr lauert. Auch Datengenerierende und Datenverarbei-

227 Adner 2017:43f.; Oliveira, Lóscio 2018:8
228 Eigene Darstellung
229 Aviatar 2022

tende können dem gleichen Unternehmen angehören, aber da es sich um zwei verschiedene Rollen handelt, sind diese Akteure hier separat dargestellt. Die Datennutzenden können gleichzeitig auch die Datenquellen gewesen sein – in diesem Fall schließt sich der Kreis der Datennutzung. Diese Akteure sind im Modell jedoch ebenfalls separat dargestellt, da auch sie unterschiedliche Rollen im Datenökosystem spielen. Somit erfüllt mein Modell alle drei zuvor aufgestellten Kriterien. Denn es sind sowohl der Datenfluss als auch die beteiligten Akteure und ihre Rollen klar erkennbar.

Die Pfeile in Abbildung 3 verdeutlichen den Datenstrom und zeigen auf, welcher Akteur die Daten an welchen anderen Akteur weiterreicht. Das Weiterreichen von Datenverarbeitenden zu Datennutzenden beinhaltet dabei das Anbieten der aufbereiteten Daten über eine digitale Plattform, die auch die Vermittlung zwischen den Datennutzenden und den Datenquellen ermöglicht. Diese Aufgabe des Anbietens der Daten über eine digitale Plattform könnte als eigenes Element bezeichnet werden, da sie ein bedeutender Schritt in der Datenvermittlung ist.[230] In meinem Modell wird das Anbieten der Daten jedoch mit dem Pfeil von Datenverarbeitenden zu Datennutzenden dargestellt und somit als Teil des Datenflusses definiert. Denn das reine Anbieten der Daten hat keine ethischen Implikationen, solange erklärt wird, wie die angebotenen Ergebnisse zustande gekommen sind. Diese Erklärung erwarte ich von den Datenverarbeitenden. Sie muss über die datenanbietende Plattform zusammen mit den Daten weitergegeben werden.

Anhand des Beispiels der Corona-Pandemie lässt sich verdeutlichen, welche Rolle Datenökosysteme für staatliches Handeln in Krisen spielen könnten. Patient:innen mit einer Corona-Erkrankung könnten Daten über ihren Krankheitsverlauf und die Wirksamkeit verschiedener Behandlungen an Ärzt:innen weitergeben – die Patient:innen würden dadurch zu Datenquellen. Die Ärzt:innen würden die Daten in ihrer Rolle als Datengenerierende an Datenverarbeitende weiterleiten. Letztere könnten mit Hilfe von Algorithmen große Mengen von Patient:innendaten verarbeiten und Informationen generieren, die sowohl durch die Patient:innen selbst, als auch Ärzt:innen oder den Staat genutzt werden könnten. Dieses Vorgehen könnte zeigen, welche Behandlungsmethoden bei wem am besten

230 Shah et al. 2020:108f.

anschlagen und sogar Aufschluss geben über die Gründe für einen schweren Krankheitsverlauf. So unterstützen Datenökosysteme die Verarbeitung und Analyse von Daten und damit das Generieren von Informationen, die für staatliches Handeln in der Krise hilfreich sein können.

Im Idealfall erfüllen alle Akteure ihre Rolle im Datenökosystem, weil sie einen Anreiz dazu haben, zum Beispiel, weil sie selbst einen Vorteil daraus ziehen. Das heißt, die Akteure werden entweder monetär entlohnt oder profitieren von aussagekräftigen Informationen, die durch ihre Teilnahme entstehen.[231] Datenquellen hätten den größten Anreiz, ihre Daten zur Verfügung zu stellen, wenn sie dafür bezahlt würden oder wenn sie selbst die verarbeiteten Daten nutzen könnten. De facto wissen Datenquellen jedoch oft nicht, dass bzw. wofür ihre Daten zusätzlich verwendet werden und können somit nicht von ihrer Rolle im Datenökosystem profitieren. Informierte Datenquellen können aber bewusst zu Datennutzenden werden und dadurch den Datenkreislauf schließen. Dabei bleiben Datenquellen und Datennutzende trotzdem zwei eigenständige Akteure, da sie unterschiedliche Motivationen haben. Der Anreiz für Datengenerierende, die Daten zu sammeln, ist genau der Gleiche – sie profitieren von Bezahlung oder den ausgewerteten Daten. Wenn zum Beispiel eine Schulleitung die Leistungsdaten ihrer Schüler:innen sammelt, um sie analysieren zu lassen, will sie auf Grundlage der verarbeiteten Daten ihr Lehrangebot verbessern.

Es ist denkbar, dass Datenverarbeitende sogar bereit wären für diese Daten zu zahlen. Denn wenn sie ein KI-System entwickelt haben, muss der Algorithmus mit vielen Daten trainiert werden, damit er seine Ergebnisse optimieren kann. Somit haben Daten (von den Quellen selbst oder von Datengenerierenden) einen großen Wert für Datenverarbeitende. Schließlich können Datenverarbeitende die Ergebnisse ihres Algorithmus an Datennutzende verkaufen und somit ihre Investition in die Daten wieder hereinholen. Wenn Datenquellen oder Datengenerierende zu Datennutzenden werden, ist der Datenkreislauf vollständig und Daten sind die einzig notwendige Währung.

231 Oliveira et al. 2019:618

3 Wissen, Kontrolle und Risiken im Datenökosystem

In diesem Kapitel wird analysiert, welche Probleme Datenökosysteme bergen. Denn die soeben beschriebene Struktur von Datenökosystemen sorgt durch die Rollenverteilung unter den Akteuren für asymmetrisches Wissen sowie für ungleiche Kontrollmöglichkeiten. Diese Tatsache sorgt dafür, dass die Akteure mit mehr Wissen und Kontrolle den Akteuren mit weniger Wissen und Kontrolle Risiken auferlegen. Da diese Risikoauferlegung ethisch relevant ist und somit eine Rolle spielt für sowohl die ethische Bewertung als auch die ethische Gestaltung von Datenökosystemen, soll in diesem Kapitel erklärt werden, warum es dazu kommt.

3.1 Epistemische Asymmetrie im Datenökosystem

In diesem Teilkapitel wird die soeben analysierte Struktur von Datenökosystemen aus einer epistemischen Perspektive betrachtet – das heißt, es wird gefragt welche Akteure was wissen über die Abläufe im Datenökosystem. Es wird sich zeigen, dass durch die Rollenverteilung im Datenökosystem den Datenquellen und Datennutzenden essentielles Wissen fehlt und dass ihnen dadurch Risiken auferlegt werden.

Dem Epistemologen Paul Humphreys zufolge ist ein Prozess für eine Person epistemisch undurchsichtig (epistemically opaque), wenn die Person nicht alle epistemisch relevanten Elemente des Prozesses kennt.[232] Das kann zum Beispiel dann der Fall sein, wenn das Wissen über den Prozess ungleich verteilt ist. Solche Asymmetrien des Wissens oder auch epistemische Asymmetrien gibt es immer dann, wenn einer der Akteure nicht über das Wissen der

232 Humphreys 2009:618

anderen Akteure verfügt. Bei dem nur einseitig vorhandenen Wissen kann es sich um spezialisiertes technisches Wissen oder Wissen über relevante organisatorische Abläufe von Fachleuten handeln, aber es kann auch in Alltagssituationen auftreten, zum Beispiel zwischen Ärzt:inen und Patient:innen.[233] Aus der Sicht des Sprachwissenschaftlers Paul Drew hängt das Wissen in jedem Fall von Rollenidentitäten ab. Außerdem werde die Asymmetrie des Wissens mit der Benachteiligung eines der Beteiligten und mit kommunikativen Schwierigkeiten in Verbindung gebracht.[234]

Eine Asymmetrie des Wissens bedeute jedoch nicht, dass der Akteur mit weniger Wissen nichts weiß. Bestimmtes Wissen wird nur Fachleuten zugeschrieben. So gelten zum Beispiel medizinische Diagnosen als normativ dem Arzt oder der Ärztin zugehörig. Die Patient:innen mögen trotzdem ihre eigenen Diagnosen kennen, aber sie haben das Wissen, weil es ihnen mitgeteilt wurde, und deshalb behandeln sie dieses Wissen nicht so, als wäre es ihr eigenes. In solchen Situationen handelt es sich auf Seiten der Patient:innen um vermitteltes Wissen und nicht um Wissen aus erster Hand. Wenn man aber Wissen aus erster Hand hat, ist man eine maßgebliche Quelle des Wissens (authoritative sources of knowledge) und hat epistemische Autorität.[235]

Laut Drew ist die Asymmetrie des Wissens eine natürliche oder notwendige Folge der ungleichen Verteilung des Wissens. Aber diese natürliche Ungleichverteilung könnte von dem Akteur mit epistemischer Autorität als strategisches Mittel eingesetzt werden, um mit dem Akteur mit weniger Wissen beliebig umzugehen – das nennt Drew eine interaktionelle (interactional) Asymmetrie. Die Möglichkeit, das asymmetrische Wissen auszunutzen, heißt jedoch nicht, dass die Asymmetrie mit Kontrolle oder Dominanz gleichzusetzen ist. Der Akteur mit vermitteltem Wissen wird also nicht notwendigerweise von dem Akteur mit epistemischer Autorität kontrolliert oder dominiert.[236]

233 Drew 1991:22, 25; Russo et al. 2024:1592
234 Drew 1991:25
235 Ariss 2009:910; Drew 1991:39f.; Snoek Henkemans, Wagemans 2012:347, Fußnote 2
236 Drew 1991:30, 32, 41, 43

Epistemische Dominanz erfordert dem Philosophen Keith Harris zufolge vielmehr die nicht wechselseitige Kompetenz eines Akteurs, die für einen anderen Akteur verfügbaren Beweise zu kontrollieren. Diese Kompetenz kann dadurch ausgeübt werden, dass Beweise vorgelegt, zurückgehalten oder gestaltet und verfügbar gemacht werden. Die Kompetenz der epistemischen Dominanz kann aber auch in Beziehungen vorhanden sein, in denen sie nicht ausgeübt wird. Der epistemisch dominante Akteur hat damit ein gewisses Maß an Kontrolle über die Fähigkeiten, das Wissen, die gerechtfertigten Überzeugungen und das Verständnis des dominierten Akteurs. Das heißt aber nicht, dass diese Kontrolle notwendigerweise bösartig genutzt wird. Die epistemische Dominanz muss zudem nicht beabsichtigt sein und kann aus der Erfüllung gesellschaftlicher Erwartungen oder anderen Zwängen entstehen.[237]

Es gibt verschiedene Abstufungen der epistemischen Dominanz: Wenn eine Person für die Wissensbildung vollkommen abhängig ist von nur einem Akteur, wie zum Beispiel ein Kind von seinen Eltern, dann ist der Einfluss dieses Akteurs sehr groß. Wenn es aber mehrere Akteure gibt, die Einfluss nehmen auf die einer Person zur Verfügung stehenden Beweise, ist jeder einzelne Akteur weniger einflussreich, und die epistemische Dominanz (epistemic domination) jeder der Akteure nimmt ab.[238]

Zudem gibt es verschiedene Möglichkeiten, epistemische Dominanz auszuüben. So können die verfügbaren Beweise eines Akteurs bei einem Ungleichgewicht durch eine Stellungnahme (testimony) direkt kontrolliert werden.[239] Ein solches Ungleichgewicht kann dadurch entstehen, dass Stellungnahmen nur einseitig möglich sind, wenn zum Beispiel Politiker:innen der Öffentlichkeit Stellungnahmen präsentieren, während die einzelnen Mitglieder der Öffentlichkeit aber nicht in der Lage sind, Politiker:innen ebenfalls mit einer Stellungnahme zu erreichen. Epistemische Dominanz kann aber auch ausgeübt werden, indem man die Stellungnahmen kontrolliert, welche der dominierte Akteur von anderen erhält. Außerdem können die Medien kontrolliert werden, zu denen der dominierte Akteur Zugang hat. Oder es kann das weitere epistemische Umfeld der

237 Harris 2022:134f.
238 Ebd.:135
239 Ebd.:135f.

dominierten Person gestaltet werden, um die Informationen zu kontrollieren, die sie erreichen. Ein Beispiel sind Eltern, die bestimmen, mit wem ihre Kinder interagieren. Als epistemische Dominanz gilt laut Harris auch die Beschränkung des Zugangs zu Informationen oder Beweisen. Zudem können gefälschte Beweise vorgelegt werden, oder es können Informationen so bearbeitet werden, dass sie ohne Kontext irreführend sind.[240]

Laut Harris ist es unvermeidlich, dass ein gewisses Maß an epistemischer Dominanz vorhanden ist. Man kann nicht all sein Wissen aus erster Hand gewinnen und muss sich deshalb auch auf vermitteltes Wissen verlassen. Wenn die dominierenden Akteure wohlmeinend und gut informiert sind, dann kann das den dominierten Akteuren epistemisch zugutekommen – sie können etwas lernen und aus der Beziehung der epistemischen Dominanz einen Vorteil ziehen. Epistemische Dominanz kann aber auch epistemisch schädlich sein für den dominierten Akteur. Dieser Schaden kann unabsichtlich verursacht werden durch einen Akteur, der selbst epistemisch dominiert wird und diese Dominanz unbeabsichtigt auf andere ausdehnt, über die er Kontrolle hat. Aber es ist auch möglich, dass dominante Akteure die Unwissenheit anderer absichtlich ausnutzen, um ihre eigene epistemische Dominanz auszuweiten. Eine Situation, in welcher der dominierte Akteur absichtlich daran gehindert wird, epistemisch wertvolle Zustände zu erreichen, nennt Harris bösartige epistemische Dominanz (*malignant epistemic domination*).[241]

Diese bösartige epistemische Dominanz kann laut Harris negative Konsequenzen haben: „If the dominating party is ill-intentioned, the dominated party is likely to form false beliefs."[242] Insbesondere, wenn der dominierende Akteur schlecht informiert ist oder schlechte Absichten hat, kann der dominierte Akteur verzerrten Ergebnissen ausgesetzt werden und somit Schwierigkeiten haben, epistemisch wertvolle Zustände zu erreichen. Außerdem können Überzeugungen, welche der dominierte Akteur bisher hatte, untergraben werden. Eine weitere Gefahr liegt darin, dass bösartige epistemische Dominanz bei dem dominierten Akteur Orientierungslosigkeit auslöst. Das heißt, der dominierte Akteur weiß nicht mehr, was er

240 Harris 2022:136
241 Ebd.:136f. 139
242 Ebd.:137

glauben soll, wem er vertrauen kann und wie diese Unsicherheit aufgelöst werden kann. Orientierungslosigkeit kann das Ergebnis eines Wettbewerbs um die epistemische Vorherrschaft sein, aber sie kann auch das ausdrückliche Ziel bestimmter Akteure sein. Im letzteren Fall kann sie durch Manipulation der epistemischen Landschaft verstärkt werden und die Evidenz selbst kann ihre Bedeutung verlieren. Orientierungslosigkeit kann aber auch schon dadurch ausgelöst werden, dass der dominierte Akteur erkennt, dass er epistemisch dominiert wird. Diese Erkenntnis erweckt Misstrauen in die eigenen Überzeugungen und auch gegenüber vermeintlich verlässlichen Instanzen wie Wissenschaftler:innen.[243]

3.1.1 Übertragung auf Datenökosysteme

In Datenökosystemen besteht eine epistemische Asymmetrie, weil Datengenerierende aus erster Hand wissen, was mit den gesammelten und gespeicherten Daten passiert und weil Datenverarbeitende aus erster Hand wissen, wie die Ergebnisse zustande gekommen sind. Datenquellen und Datennutzende hingegen sind darauf angewiesen, auf das durch Datengenerierende und Datenverarbeitende vermittelte Wissen zurückgreifen. Dadurch, dass Datengenerierende und Datenverarbeitende Wissen aus erster Hand haben, sind sie eine maßgebliche Quelle des Wissens und haben epistemische Autorität.[244] Die Akteure mit epistemischer Autorität können die Ungleichverteilung von Wissen als strategisches Mittel einsetzen, um mit dem Akteur mit weniger Wissen beliebig zu verfahren.[245] Es ist also möglich, dass Datengenerierende und Datenverarbeitende ihr Wissen gegenüber Datenquellen und Datennutzenden als strategisches Mittel einsetzen, dennoch werden Letztere nicht automatisch von Ersteren kontrolliert oder dominiert.[246]

Um epistemische Dominanz handelt es sich aber dann, wenn ein Akteur dazu fähig ist, die für einen anderen Akteur verfügbaren Beweise zu kontrollieren und damit zu beeinflussen, zu welchen

243 Harris 2022:134, 137–139
244 Drew 1991:39f.
245 Ebd.:32
246 Ebd.:41

Überzeugungen der andere Akteur kommt.[247] Datengenerierende sind zu einem gewissen Grad dazu in der Lage, die für Datenquellen verfügbaren Beweise in Bezug darauf, was mit den Daten passiert, zu kontrollieren. Und Datenverarbeitende sind zu einem gewissen Grad dazu in der Lage, die für Datennutzende verfügbaren Beweise in Bezug darauf, wie das Ergebnis zustande gekommen ist, zu kontrollieren. Datenquellen und Datennutzende sind dennoch in Bezug auf ihre Daten und die algorithmischen Ergebnisse nicht vollkommen epistemisch dominiert, da ihre Wissensbildung nicht ausschließlich abhängig ist von jeweils nur dem einen Akteur. Denn Datenquellen können auch an anderer Stelle zumindest allgemeine Informationen darüber bekommen, was passieren kann, wenn personenbezogene Daten zentral gespeichert werden und Datennutzende können ebenfalls aus anderen Quellen allgemeine Informationen darüber beziehen, was passieren kann, wenn Daten von Algorithmen analysiert werden.

Aber in Bezug auf das konkrete Datenökosystem wissen nur die Datengenerierenden, wie genau die Daten gespeichert werden und welche Maßnahmen ergriffen wurden, um der Zweckentfremdung durch Unbefugte vorzubeugen – nur dieser eine Akteur weiß, wie sicher die personenbezogenen Daten der Datenquellen sind und zu welchen Zwecken sie verarbeitet werden sollen. Die Informationen darüber, welche Daten und welcher Algorithmus genutzt wurden, um das Ergebnis zu generieren, haben nur die Datenverarbeitenden. Durch diese epistemische Dominanz können Datengenerierende und Datenverarbeitende die verfügbaren Beweise der Datenquellen und Datennutzenden durch ihre Stellungnahmen direkt kontrollieren und könnten zum Beispiel auch den Zugang zu Informationen beschränken.[248] Das heißt aber noch nicht, dass diese epistemische Dominanz auch bösartig eingesetzt wird. Solange die dominierenden Akteure wohlmeinend und gut informiert sind, kann das den dominierten Akteuren epistemisch zugutekommen – in dem Fall erhalten Datenquellen und Datennutzende alle Informationen, die sie brauchen um zu epistemisch wertvollen Überzeugungen zu gelangen. Bei schlechten Absichten der Datengenerierenden oder Datenverarbeitenden kann diese epistemische Dominanz schädlich sein

247 Harris 2022:134, 137
248 Ebd.:135f.

für Datenquellen oder Datennutzende – zum Beispiel, wenn die Sicherheit der Daten beschönigt wird, um möglichst viele Daten zu sammeln. Ein solcher Schaden kann aber auch unabsichtlich dadurch verursacht werden, dass Datengenerierenden oder Datenverarbeitenden es selbst nicht besser wissen.[249]

Die Prozesse in Datenökosystemen sind in einem solchen Fall für Datenquellen und Datennutzende epistemisch undurchsichtig (epistemically opaque), weil sie nicht alle epistemisch relevanten Elemente des Prozesses kennen.[250] Denn die epistemisch relevanten Elemente des Prozesses bei der Datenverarbeitung im Datenökosystem beziehen sich auf Wissen über den Datenfluss. Auf Grund der epistemischen Asymmetrie besteht aber die Gefahr, dass Datenquellen nicht darüber informiert werden, zu welchem Zweck ihre Daten verarbeitet werden, und Datennutzenden nicht wissen, wie die Ergebnisse, die ihnen zur Verfügung gestellt werden, zustande gekommen sind.

Datengenerierende und Datenverarbeitende haben zwar Wissen aus erster Hand über die epistemisch relevanten Elemente des Prozesses bezüglich der Nutzung und Verarbeitung der Daten im Datenökosystem, aber sie haben kein vollkommenes Wissen darüber. Ob personenbezogene Daten zweckentfremdet werden, kann nur mit einer gewissen Wahrscheinlichkeit ausgeschlossen werden, und auch ein Datendiebstahl kann den Verantwortlichen vorerst verborgen bleiben. Die Entstehung der Ergebnisse von Datenverarbeitung ist ebenfalls nicht vollkommen durchsichtig, wenn komplexe KI-Systeme eingesetzt werden.[251] Dann können selbst die Akteure mit mehr Wissen nicht mehr nachvollziehen, wie genau die Ergebnisse zustande gekommen sind, und es besteht ein Risiko, dass sie verzerrt sind.

Dieses Risiko können dann selbst Expert:innen nicht perfekt einschätzen, da es immer eine epistemische Unsicherheit gibt.[252] Wenn zudem die Einschätzung des Risikos von denjenigen Akteuren kommt, die das Risiko den Datenquellen und Datennutzenden auferlegen, kann es zu einem Interessenskonflikt kommen. Luhmann formuliert das so: „Diejenigen, für die das Risikoverhalten anderer

249 Harris 2022:137–139
250 Humphreys 2009:618
251 Russel, Norvig 2021:801
252 Möller 2009:219f.

zur Gefahr wird, werden anders darüber urteilen als die, die die Entscheidung selbst treffen oder an ihr beteiligt sind."[253] Die Einstellung der Akteure, die anderen ein Risiko auferlegen, widerspricht aber nicht automatisch der der Risikoexponierten. Denn Erstere könnten entweder ein Interesse daran haben, dass das Risiko als gering eingeschätzt und ohne Protest getragen wird oder aber daran, das durch ihre Arbeit entstehende Risiko so genau wie möglich einzuschätzen, um nicht für mögliche Folgeschäden aufkommen zu müssen.

Selbst wenn Datengenerierende und Datenverarbeitende das durch ihre Arbeit entstehende Risiko so gut wie möglich einschätzen wollen, bleibt immer eine Restunsicherheit.[254] Denn die risikoauferlegenden Akteure haben ja nicht die volle Kontrolle über die Risiken, die sie verursachen. So können Datengenerierende zum Beispiel personenbezogene Daten sammeln und so sicher wie möglich speichern, doch es ist schwer vorauszusehen, wie groß das Interesse an den Daten ist, und mit welchen Methoden womöglich Angreifer versuchen könnten, die Sicherheitsvorkehrungen zu umgehen. Gleichermaßen können Datenverarbeitende sich größte Mühe geben, korrekte algorithmisch generierte Ergebnisse zur Verfügung zu stellen, aber insbesondere, wenn KI-Systeme eingesetzt werden, ist das Risiko verzerrter Ergebnisse nicht leicht einzuschätzen.

Es besteht bei allen Akteuren epistemische Unsicherheit über die Folgen des Datenökosystems. Aber nur Datenquellen und Datennutzende tragen das Risiko, dass ihre Daten ohne ihr Wissen zweckentfremdet werden, oder dass ihnen verzerrte Ergebnisse zur Verfügung gestellt werden, auf deren Grundlage sie zu falschen Überzeugungen gelangen. Deshalb argumentiere ich, dass Datenquellen auf Grund der Struktur eines Datenökosystems epistemisch benachteiligt sind, da sie nicht über die Verwendung ihrer personenbezogenen Daten Bescheid wissen und damit für sie das Risiko der Verletzung der Privatsphäre besteht. Für Datennutzende besteht auf Grund der epistemischen Asymmetrie die Möglichkeit, einem verzerrten Ergebnis zu glauben, weshalb fehlende Transparenz für sie ein Risiko birgt.

253 Luhmann 1990:163
254 Ebd.

3.1.2 Risiko: Definition

Der Begriff „Risiko" kann unterschiedliche Bedeutungen haben.[255] Nach Ansicht des theoretischen Philosophen Sven Ove Hansson bezieht sich der Begriff Risiko auf Situationen, in denen es möglich, aber nicht sicher ist, dass ein unerwünschtes Ereignis eintreten wird. Mit Risiko kann aber auch die Ursache, die Wahrscheinlichkeit oder der statistische Erwartungswert (Schadenhöhe mal Eintrittswahrscheinlichkeit) eines unerwünschten Ereignisses gemeint sein, welches eintreten oder nicht eintreten kann. Zudem wird Risiko im Rahmen der Entscheidungstheorie interpretiert als die Tatsache, dass eine Entscheidung unter den Bedingungen bekannter Wahrscheinlichkeiten getroffen wird.[256] Der Soziologe Niklas Luhmann hingegen meint: „Üblicherweise spricht man von Risiko immer dann, wenn ein möglicher Schaden um eines Vorteils willen in Kauf genommen wird."[257] Seiner Meinung nach können Risiken dementsprechend durch Verzicht auf die Vorteile vermieden werden.

Auf eins können sich jedoch laut Luhmann alle Definitionen einigen: „(...) der Risikobegriff verweist auf Zukunft."[258] Fest steht, dass das in den Definitionen betrachtete unerwünschte Ereignis, beziehungsweise der mögliche Schaden, immer in der Zukunft liegt. Dem Wirtschaftswissenschaftler Frank H. Knight zufolge unterscheidet die Fähigkeit, sich auf eine Situation einstellen zu können, bevor diese Situation eintritt, uns Menschen von Tieren. Je weiter man vorausschauen könne, desto besser könne man sich anpassen und desto kompetenter könne man leben.[259] Der Medizinethiker Joschka Haltaufderheide meint, dass dies das menschliche Zeitverständnis in der Neuzeit prägt: „Wir leben und handeln in der Erwartung und auf das hin, was noch nicht ist. Die Kenntnis der Zukunft, zumindest ihrer Struktur nach, ist daher essentiell, um eine Antwort auf die Frage formulieren zu können, was zu tun ist."[260] Somit sei auch jegliche ethische Einschätzung abhängig von dem Blick auf potenziell eintretende unerwünschte Ereignisse.

255 Hansson 2005:7f.
256 Hansson 2004:10
257 Luhmann 1990:135
258 Ebd.:140
259 Knight 1921:200
260 Haltaufderheide 2015:14

Der Technologiephilosoph Günter Ropohl charakterisiert den Begriff des Risikos anhand des Schadens, den ein unerwünschtes Ereignis verursacht, und anhand der Eintrittswahrscheinlichkeit dieses unerwünschten Ereignisses. Dabei unterschiedet er verschiedene Formen des Risikos anhand von folgenden Merkmalen: Der Schaden, der riskiert wird, kann unterschiedlich hoch sein, und diese Schadenshöhe kann unterschiedlich leicht bestimmbar sein. Zudem variiert, inwiefern die Schadenshöhe beeinflussbar ist durch zum Beispiel Vorkehrungen in Form von Regeln. Die Ursache des Schadens könne natürlichen Ursprungs sein oder durch Menschen verursacht werden – durch die Risikoexponierten selbst oder durch Außenstehende. Die Bekanntheit des Schadenspotenzials spielt laut Ropohl eine Rolle bei seiner Akzeptanz. Auch die Eintrittswahrscheinlichkeit des Risikos ist mit unterschiedlicher Gewissheit bestimmbar und manchmal durch die Betroffenen, durch Außenstehende oder gar nicht beeinflussbar.[261]

Luhmann würde Ropohl in dem Punkt widersprechen, dass der Schaden, der riskiert wird, auch natürlichen Ursprungs oder durch Außenstehende verursacht sein kann. Denn er ist der Meinung, dass man sich nur selbst einem Risiko aussetzen kann. Wenn Außenstehende einen möglichen Schaden „(...) um eines Vorteils willen in Kauf [nehmen] (...)“[262], dann sei man hingegen einer Gefahr ausgesetzt. Das Wesentliche bei seiner begrifflichen Unterscheidung ist die Möglichkeit, selbst eine Entscheidung zu treffen: Trifft man eine Entscheidung, die potenziell zu zukünftigen Schäden führt, „(...) geht man mit der Entscheidung ein Risiko ein. Schäden, die außerhalb dieses Einflußbereiches liegen, werden, solange sie noch unsicher sind, als Gefahr angesehen.“[263] So seien auch natürlich verursachte Katastrophen keine Risiken, sondern Gefahren. Hat man jedoch Entscheidungsmöglichkeiten, „die einen etwaigen Schadenseintritt beeinflussen bzw. ihn vermeiden helfen könnten.“[264], so kann man laut Luhmann wieder von einem Risiko sprechen, wenn die Möglichkeiten bewusst nicht ergriffen werden.

261 Ropohl 2016:12ff.
262 Luhmann 1990:135
263 Ebd.:149
264 Ebd.:150

Luhmann unterscheidet insbesondere deshalb zwischen Risiko und Gefahr, weil es Situationen gibt, „(...) in denen das Risikoverhalten des einen zur Gefahr für den anderen wird."[265] Dabei ist es möglich, dass das Risiko zwar rational kalkuliert wurde, aber für andere, die an der Entscheidung nicht beteiligt waren, die jedoch von den potenziellen Schäden zukünftig betroffen sein könnten, daraus eine Gefahr entsteht: „Das riskante Verhalten der einen wird zur Gefahr für die anderen (...)."[266] Luhmann behauptet, dass man in Gefahr ist, wenn potenzieller Schaden durch „(...) Ursachen außerhalb der eigenen Kontrolle (...)"[267] entsteht, also von Außenstehenden oder natürlichen Umständen verursacht wird. Während man sich gegenüber natürlichen Gefahren entweder risikoavers oder risikoaffin verhalten könne, sei man den Gefahren, die durch die Entscheidungen von Außenstehenden entstehen, hilflos ausgesetzt.[268] Auch risikoaffine Menschen seien nicht unbedingt bereit dazu, diese, durch riskante Entscheidungen von Außenstehenden verursachten Gefahren, hinzunehmen.[269] Dass wir Menschen aber im Alltag nicht umhin können, das von Außenstehenden verursachte Risiko hinzunehmen (zum Beispiel im Straßenverkehr), wird in Kapitel 5 dargestellt.

Betrachten wir nun die Perspektive der Außenstehenden, welche durch ihre Entscheidungen Risiken entstehen lassen. Die Entscheidungstheorie geht davon aus, dass den Entscheidenden bekannt ist, welche Ergebnisse möglich sind und mit welcher Wahrscheinlichkeit sie eintreten werden.[270] Sicher bekannt sind die Wahrscheinlichkeiten der möglichen Ergebnisse allerdings nur in idealisierten Fallbeispielen. So schreibt Hansson: „Only very rarely are probabilities known with certainty."[271] Denn in der Realität ist es laut Haltaufderheide oft nicht möglich, „(...) sich der vollständigen Bekanntheit aller kausalen Einflüsse auf ein zukünftiges Ereignis sicher zu sein (...)."[272] Deshalb handelt es sich laut Hansson meist tatsächlich um Entscheidungen ‚unter Unsicherheit': „In decision-making under un-

265 Luhmann 1990:152
266 Ebd.:161
267 Ebd.:149
268 Ebd.:153
269 Ebd.:154
270 Hansson 1999:539
271 Ebd.
272 Haltaufderheide 2015:153

certainty, probabilities are either not known at all or are only known with insufficient precision."[273] Dann gibt es keine Möglichkeit, die Auswirkungen einer Entscheidung abzuschätzen.

Um in dem von Unsicherheiten geprägten Alltag Entscheidungen treffen zu können, müssen wir Menschen Unsicherheiten reduzieren und Überzeugungen fixieren. Das heißt, Aussagen über die Welt werden nicht nur Wahrscheinlichkeiten zugewiesen, sondern sie werden entweder für wahr oder falsch gehalten. Auf diese Weise ist die unsichere Welt kognitiv handhabbar und Entscheidungen können leichter getroffen werden. Eine solche Reduzierung der Komplexität darf allerdings nur vorrübergehend sein, denn wenn es Gründe dafür gibt, müssen wir von der Überzeugung zur Wahrscheinlichkeit oder sogar zur Unsicherheit zurückkehren können.[274]

Wenn also Entscheidungen ‚unter Risiko' getroffen werden, ist nicht tatsächlich vollständig bekannt, mit welcher Wahrscheinlichkeit unerwünschte Ereignisse eintreten. Vielmehr wird die Beschreibung der Entscheidungsprobleme vereinfacht, indem sie als Fälle mit bekannten Wahrscheinlichkeiten behandelt werden.[275] Laut Luhmann transformiert die Wahrscheinlichkeitsrechnung Unsicherheit in sichere Entscheidungen.[276]

Das Gegenteil von Risiko ist jedoch laut dem praktischen Philosophen Niklas Möller, dem theoretischen Philosophen Sven Ove Hansson und dem Ethiker Martin Peterson nicht absolute Sicherheit. Sie sind der Meinung, dass die Annahme je geringer das Risiko, desto höher die Sicherheit auf einer unzureichenden Definition von Sicherheit basiert. Um diese Meinung zu begründen, unterscheiden Möller et al. zwischen einem absoluten und einem relativen Sicherheitskonzept. Absolute Sicherheit bedeutet demnach, dass das Risiko eines Schadens vollständig beseitigt ist, während relative Sicherheit bedeutet, dass das Risiko auf ein bestimmtes (tolerierbares) Niveau reduziert wurde. Die Autoren sind sich einig, dass das Konzept von absoluter Sicherheit nur ein Ideal ist, das niemals erreicht werden kann. Auch Luhmann stimmt zu: „Es gibt aber keine risikofreie

273 Hansson 1999:539
274 Hansson 2004:12–15; Hansson 1999:540
275 Hansson 1999:539; Hansson 2003:293; Hansson 2004:11f.
276 Luhmann 1990:136

Sicherheit."[277] Vielmehr lassen Sicherheitskonzepte in der Realität meist Risikowahrscheinlichkeiten über Null zu. Gemäß einem solchen relativen Sicherheitskonzept bedeutet also die Aussage „diese Daten sind sicher" das gleiche wie „die Sicherheit der Daten ist so hoch, wie es unter Berücksichtigung angemessener Kosten und vorbeugender Maßnahmen erwartet werden kann, und das Risiko einer Zweckentfremdung der Daten ist sehr gering".[278]

Sicherheit ist aber auch aus einem weiteren Grund nicht unbedingt immer der Gegenspieler von Risiko. Denn der Schaden, den ein Risiko auslöst, kann laut Luhmann „(...) auch im Ausbleiben eines Vorteils [bestehen], in dessen Erwartung man investiert hatte."[279] Zum Beispiel besteht bei der Teilnahme an einer Lotterie das Risiko, dass man eine Niete zieht – dies wäre ein unerwünschtes Ereignis. Die Vermeidung dieses Ereignisses ist aber keine Frage der Sicherheit.[280] Sicherheit kann nur dann im Angesicht eines Risikos erstrebenswert sein, wenn das unerwünschte Ereignis einen Schaden verursacht und damit das Gegenteil von Sicherheit erzeugt.

Ich gehe im Folgenden davon aus, dass ein Risiko eine Situation beschreibt, in welcher ein unerwünschtes Ereignis mit einer gewissen Wahrscheinlichkeit eintritt. In solche Situationen kann sich ein Individuum selbst bringen und sich damit selbst einem Risiko aussetzen. Aber Risiken können einem Individuum auch von Außenstehenden auferlegt werden – dadurch entsteht für das Individuum die Gefahr, dass das unerwünschte Ereignis tatsächlich eintritt. In erster Linie wird das Individuum aber, durch die Entscheidungen von Außenstehenden, in eine Situation gebracht, in welcher unsicher ist, ob das unerwünschte Ereignis eintritt oder nicht.

Datenquellen und Datennutzende wissen auf Grund der epistemischen Asymmetrie im Datenökosystem nicht Bescheid, wie mit den sie betreffenden Daten umgegangen wird. So fehlt Datenquellen das Wissen darüber, zu welchem Zweck ihre Daten verarbeitet werden und Datennutzenden fehlt das Wissen über die Entstehung des Ergebnisses, dem sie ausgesetzt sind. Das heißt, im Datenökosystem können zwei unerwünschte Ereignisse mit einer gewissen Wahrscheinlichkeit eintreten – es entstehen Risiken: Für Datenquellen

277 Luhmann 1990:136
278 Möller et al. 2006:420
279 Luhmann 1990:138
280 Möller et al. 2006:421

besteht das Risiko, dass ihre Daten zu Zwecken verwendet werden, denen sie nicht zugestimmt haben oder dass sogar Unbefugte Zugriff auf ihre Daten bekommen. Für Datennutzende besteht das Risiko einem verzerrten Ergebnis zu glauben und somit zu falschen Überzeugungen zu gelangen, weil sie die Entstehung nicht nachvollziehen können. Im nächsten Teilkapitel soll gezeigt werden, dass Datenquellen und Datennutzende neben dem epistemischen Nachteil auch weniger Kontrolle haben über die Vorgänge im Datenökosystem und dass ihnen dadurch die Risiken auferlegt werden.

3.2 Verteilung der Kontrollmöglichkeiten im Datenökosystem

Wie in Teilkapitel 3.1 festgestellt, haben die epistemisch dominanten Akteure ein gewisses Maß an Kontrolle über die Fähigkeiten, das Wissen, die gerechtfertigten Überzeugungen und das Verständnis der epistemisch dominierten Akteure.[281] In diesem Teilkapitel sollen die Kontrollmöglichkeiten der Akteure über die Vorgänge im Datenökosystem genauer betrachtet werden. Zur Veranschaulichung möchte ich nun mein Modell der ethisch relevanten Akteure eines Datenökosystems auf die öffentliche Verwaltung anwenden und den hypothetischen Fall eines digitalisierten Wohngeldantrages betrachten.

Im Falle der öffentlichen Verwaltung ist das Ziel der Datenverarbeitung im Datenökosystem eine effektivere Bearbeitung von Anträgen, z.B. auf Wohngeld.[282] Bei der digitalen Gestaltung des Wohngeldantrages in Form eines Datenökosystems, fließen die Daten der Antragstellenden (Datenquellen) zur Wohngeldbehörde. Diese sammelt die Daten ein und nimmt damit die Rolle als Datengenerierende ein. Sie leitet die Daten weiter an das Statistische Landesamt, an das Statistische Bundesamt sowie an das Bundesministerium des Innern, für Bau und Heimat und an das Bundesamt für Bauwesen und Raumordnung, damit diese die Daten im Rahmen einer Wohngeldstatistik verarbeiten können.[283] Gleichzeitig ist die Wohngeldbehörde auch für die Verarbeitung der Daten zuständig, da sie den Antragstellenden eine Rück-

281 Harris 2022:134f.
282 §1a, Abs. 1, Satz 1, OZG
283 §§ 34–36 WoGG

meldung über deren Anspruch auf Wohngeld und dessen Höhe geben muss – sie repräsentiert somit auch die Akteursgruppe der Datenverarbeitenden. Datennutzende sind dann sowohl die Antragstellenden, die erfahren, ob sie Wohngeld bekommen, als auch der Staat, der anhand der Wohngeldstatistik erkennen kann, an welchen Orten in Deutschland bezahlbare Wohnungen fehlen.

Von einem solchen Datenökosystem profitieren sowohl Antragstellende (Datenquellen und Datennutzende), die Wohngeldbehörde (Datengenerierende und -verarbeitende) als auch der Staat (ebenfalls Datennutzende). Die Antragstellenden müssen nicht mit all ihren Dokumenten zum Amt, sondern sie können schnell und von Zuhause aus erfahren, ob sie einen Anspruch auf Wohngeld haben. Die Wohngeldbehörde hat einen geringeren Arbeitsaufwand, wenn sie alle notwendigen Dokumente digital abrufen kann und bei der Einschätzung des Anspruchs durch einen Algorithmus unterstützt wird. Der Staat profitiert von der Weitergabe und Auswertung der Daten sowie einer kürzeren Bearbeitungszeit.

Denn wir nehmen an, dass ohne schnellere Antragsbearbeitung mehrere Risiken bestehen: Auf Grund des Personalmangels in den Wohngeldbehörden können nicht alle Wohngeldanträge zeitnah bearbeitet werden – so müssen Personen mit Anspruch auf Wohngeld länger ohne dieses auskommen. Den Antragstellenden droht also der Wohnungsverlust, falls die Miete ohne Unterstützung nicht mehr bezahlt werden kann. Den Mitarbeitenden in den Wohngeldbehörden droht berufliche Überlastung und dem Staat fehlen aussagekräftige Wohngeldstatistiken und somit wertvolle Planungsgrundlagen für den öffentlichen Wohnungsbau.

Aber es bestehen auch Risiken durch die Datenverarbeitung im Datenökosystem. Um das Ziel der schnelleren Antragsbearbeitung erreichen zu können, ist es notwendig, die Dokumente der Antragstellenden digital zu speichern und zwischen den verschiedenen Behörden auszutauschen. Viele dieser Dokumente enthalten sensible Informationen über die Antragstellenden.[284]

Durch das digitale Speichern werden die Antragstellenden dem Risiko des Zugriffs durch Unbefugte, z.B. Hacker, und damit der

284 Zum Beispiel: Personalien, Adresse, Verdienstbescheinigung (auszufüllen vom Arbeitgeber), Einkommensnachweise (Ausbildungs- oder Arbeitsvertrag, Lohnabrechnungen), Schwerbehindertenausweis, Bescheide über Kindergeld/Kinderzuschlag/Elterngeld

Zweckentfremdung ihrer sensiblen Daten ausgesetzt. Das heißt, es kann potenziell ein Schaden auftreten. Der Schaden wird nicht zwangsläufig verursacht, aber es besteht ein Risiko. Und zum schnelleren Bearbeiten der Wohngeldanträge benötigt man noch ein weiteres Mittel: Den Algorithmus, der die Daten analysiert und berechnet, ob ein Anspruch auf Wohngeld besteht. Insbesondere wenn der Regelsatz des Algorithmus nicht durchschaubar und dadurch eventuell unbemerkt falsch ist, entsteht ein weiteres Risiko.

Es fällt auf, dass von den Risiken insbesondere die Antragstellenden in ihrer Rolle als Datenquellen und als Datennutzende betroffen sind. Datengenerierende haben die Kontrolle darüber, wie die Daten der Datenquellen gespeichert werden und an wen sie weitergegeben werden. Die Datenquellen hingegen müssen im Fall des Wohngeldantrages der Weitergabe ihrer Daten an unter anderem das Statistische Bundesamt zustimmen, um den Antrag überhaupt stellen zu können. Ebenso haben Datenverarbeitende die Kontrolle darüber, wie die Daten verarbeitet werden und mit welcher Erklärung die Ergebnisse zur Verfügung gestellt werden. So wissen die Antragstellenden in ihrer Rolle als Datennutzende nicht, mit welchem Algorithmus ihr Antrag bearbeitet wurde und können nicht überprüfen, ob dieser Algorithmus ein fehlerhaftes Regelwerk erlernt oder vorgegeben bekommen hat.

3.2.1 Drei-Parteien-Modell

Um besser zu verstehen, welche Rollen die Akteure in Bezug auf das Risiko spielen und wieviel Kontrolle sie jeweils darüber haben, wird nun das Drei-Parteien-Modell von der Technologiephilosophin und Ethikerin Hélène Hermansson und dem bereits erwähnten theoretischen Philosophen Sven Ove Hansson vorgestellt. Dieses Modell soll eine systematische Beschreibung der ethischen Aspekte des Risikos ermöglichen.[285] Die Autor:innen identifizieren in Situationen, in denen Risiken entstehen, drei Parteien: die Risikoexponierten, die Entscheidungstragenden und die durch das Risiko Begünstigten. Durch das Betrachten der Beziehungen zwischen diesen drei Parteien werden die ethischen Aspekte von Risiken identifiziert. Um diese

285 Hermansson, Hansson 2007:130

Beziehungen analysieren zu können, entwickeln die Autor:innen sieben Fragen. Diese Fragen betrachten das Verhältnis zwischen den durch das Risiko Begünstigten und den Risikoexponierten und die Verteilung von Risiken und Nutzen zwischen ihnen – insbesondere, ob diese Verteilung durch Umverteilung oder Ausgleich weniger ungerecht gestaltet werden kann.

Die Fragen betrachten außerdem die Beziehung zwischen den Risikoexponierten und den Entscheidungstragenden und den Einfluss, den die Risikoexponierten auf die Entscheidungsfindung haben. Der Einfluss der Risikoexponierten ist abhängig von den Informationen, die ihnen zur Verfügung gestellt werden: „(…) the most common ethical problem in relation to information is the degree to which experts and decision-makers provide the risk-exposed with adequate and comprehensible information about the risk (…)."[286] Zuletzt betrachten die Fragen auch die Beziehung zwischen den Entscheidungstragenden und den durch das Risiko Begünstigten. Denn wenn derselbe Akteur zu beiden Gruppen gehören sollte, stellt das seine Legitimität in Frage: „A person who will herself reap the benefits if someone else is exposed to a risk does not have much credibility as a judge of whether the benefits outweigh the risks."[287] Sollten die Entscheidungstragenden selbst von dem Risiko profitieren können, so ist die Gefahr groß, dass sie aus Eigennutz den Risikoexponierten das Risiko auferlegen.

Eine Möglichkeit ein solches eigennütziges Verhalten zu vermeiden ist, die Entscheidung, ob das Risiko eingegangen wird, den Risikoexponierten zu überlassen – so können sie das Risiko frei wählen: „(…) for a risk to be freely chosen, the risk-exposed has to have full information about the risk and has to be capable of making the decision whether or not to accept the risk."[288] So lässt sich unterscheiden, ob die Risikoexponierten das Risiko akzeptiert haben, um einen Vorteil zu erlangen, den sie sonst nicht hätten erlangen können, oder ob sie gegen ihren Willen dem Risiko ausgesetzt wurden.[289]

286 Hermansson, Hansson 2007:132
287 Ebd.
288 Ebd.:142
289 Ebd.:130

3.2.2 Anwendung des Drei-Parteien-Modells auf das Wohngeldantrag-Beispiel

Die Anwendung des Drei-Parteien-Modells ermöglicht die Analyse der ethischen Aspekte von (un-)zulässiger Auferlegung von Risiken am Beispiel des Wohngeldantrages. Zuerst müssen die drei Parteien identifiziert werden: Die Risikoexponierten sind die Antragstellenden, da ihre Daten potenziell zweckentfremdet werden können bzw. der zur Antragbearbeitung genutzte Algorithmus ihren Antrag fälschlicherweise ablehnen könnte. Die Entscheidungstragenden sind die Wohngeldbehörden oder im weitesten Sinne der Staat, welcher sich für die Digitalisierung der Wohngeldanträge entscheidet. Die durch das Risiko Begünstigten sind sowohl Antragstellende, die Wohngeldbehörde als auch der Staat. Dabei sind die Entscheidungstragenden im Kontext vom Datenökosystem die Akteure mit mehr Kontrolle und Risikoexponierte sind meist die Akteure mit weniger Kontrolle (außer Entscheidungstragende legen sich selbst ein Risiko auf).

Zuerst wird das Verhältnis zwischen den Risikoexponierten und den durch das Risiko Begünstigten betrachtet – das ist in diesem Fall das Verhältnis zwischen den Antragstellenden einerseits und der Gruppe der Antragstellenden, der Wohngeldbehörde und dem Staat andererseits. Die Verteilung von Risiken und Nutzen zwischen ihnen ist klar, denn die Risiken tragen allein die Antragstellenden, da sie nicht die Kontrolle über den Einsatz des Algorithmus haben. Der Nutzen der Digitalisierung der Wohnanträge scheint jedoch gleichmäßig verteilt zu sein – jede beteiligte Partei profitiert auf ihre Weise. Man könnte es als Ausgleich bezeichnen, dass die Antragstellenden als Risikoexponierte auch gleichzeitig Teil der Partei der durch das Risiko Begünstigten sind.

Betrachten wir die Beziehung zwischen den Risikoexponierten und den Entscheidungstragenden so fällt auf, dass sich die Akteure in diesen Parteien nicht mehr überschneiden. Die Antragstellenden sind als Risikoexponierte nicht Teil der Entscheidungstragenden, welche durch die Wohngeldbehörde und den Staat vertreten sind. Die Antragstellenden entscheiden zwar, ob sie der digitalen Sammlung ihrer Daten zustimmen, sind aber dazu gezwungen, wenn sie einen Wohngeldantrag stellen wollen. Wie sicher die Daten ausgetauscht und gespeichert werden und ob ein KI-System genutzt wird,

entscheiden die Wohngeldbehörden und der Staat. Selbst wenn die Antragstellenden über die Risiken der digitalen Datenspeicherung und der Nutzung eines Algorithmus zur Bearbeitung der Anträge informiert würden, wären sie immer noch darauf angewiesen, diese Risiken einzugehen, um sich die finanzielle Unterstützung holen zu können, die ihnen zusteht.

Bei der Betrachtung der Beziehung zwischen den Entscheidungstragenden und den durch das Risiko Begünstigten fällt auf, dass die Wohngeldbehörden und der Staat zu beiden Parteien gehören. Hier zeigt sich, dass die Wohngeldbehörden und der Staat davon profitieren, dass die Antragstellenden einem Risiko ausgesetzt werden. Dies ließe befürchten, dass die Wohngeldbehörden und der Staat den Antragstellenden das Risiko aus Eigennutz auferlegen. Da allerdings auch die risikoexponierten Antragstellenden von der Situation profitieren können, kann die Befürchtung widerlegt werden. Das Risiko wird den Antragstellenden auferlegt, damit sie selbst, sowie alle weiteren beteiligten Parteien, davon profitieren können.

Die ausgeglichene Verteilung des Nutzens zwischen Risikoexponierten und Begünstigten zeigt auf, dass den risikoexponierten Antragstellenden das Risiko nicht aus Eigennutz, sondern zur Erreichung eines Allgemeinwohls auferlegt wird. Jedoch ist die Verteilung von Risiken zwischen Risikoexponierten und Begünstigten noch immer sehr ungleich. Nicht nur die Risiken sind ungleich verteilt, sondern auch die Möglichkeit, die Risiken zu kontrollieren. Denn die Informationen, die den risikoexponierten Antragstellenden zur Verfügung gestellt werden, verleihen ihnen nicht die Kontrolle über ihre Daten oder den Verlauf der Antragsbearbeitung. Diejenigen, die das Risiko eingehen, können dieses Risiko also nicht beeinflussen.

Die Entscheidung, ob das Risiko eingegangen wird, wird nicht den Risikoexponierten überlassen – sie können das Risiko nicht frei wählen: „(…) for a risk to be freely chosen, the risk-exposed has to have full information about the risk and has to be capable of making the decision whether or not to accept the risk."[290] Die risikoexponierten Antragstellenden können zwar über die Risiken informiert werden, sie haben deshalb aber trotzdem keine gute Alternative als das Risiko zu akzeptieren. Wer einen Wohngeldantrag stellen will, um ihm zustehende staatliche Leistungen ausgezahlt zu bekommen,

290 Hermansson, Hansson 2007:142

der muss die Risiken in Kauf nehmen. Antragstellende werden dabei nicht gegen ihren Willen dem Risiko ausgesetzt – sie akzeptieren das Risiko, um einen Vorteil zu erlangen, den sie sonst nicht hätten erlangen können.[291] Das kann in dem Fall dieses Beispiels allerdings nicht als freie Wahl bezeichnet werden.[292]

Die Risiken im Datenökosystem wurden zwar durch Datengenerierende und Datenverarbeitende rational kalkuliert, aber für Datenquellen und Datennutzende, die an der Entscheidung nicht beteiligt waren, die jedoch von den potenziellen Schäden zukünftig betroffen sein könnten, entsteht daraus eine Gefahr.[293] Für Datenquellen besteht die Gefahr, dass ihre Daten ohne ihr Wissen zweckentfremdet werden, und für und Datennutzende besteht die Gefahr, dass ihnen verzerrte Ergebnisse zur Verfügung gestellt werden, auf deren Grundlage sie zu falschen Überzeugungen gelangen. In Luhmanns Worten sind Datenquellen und Datennutzende deshalb in Gefahr, weil der potenzielle Schaden durch „(...) Ursachen außerhalb der eigenen Kontrolle (...)"[294] entsteht, also von Außenstehenden verursacht wird. Dieser Gefahr, die durch die Entscheidungen von Außenstehenden entstanden ist, seien Datenquellen und Datennutzende hilflos ausgesetzt.[295]

Deshalb können Datenökosysteme verstanden werden als ein Zusammenspiel von Akteuren, in welchem die einen den anderen ein Risiko auferlegen. Datengenerierende haben nicht nur mehr Wissen, sondern auch mehr Kontrolle als die Datenquellen darüber, was mit den Daten passiert, und Datenverarbeitende haben mehr Wissen und Kontrolle darüber, welche Ergebnisse generiert werden als die Datennutzenden. Die Risiken, die dabei entstehen, wie zum Beispiel Zweckentfremdung von personenbezogenen Daten oder missverständliche algorithmische Ergebnisse, werden also durch die Akteure mit mehr Kontrolle verursacht – die Risiken werden allerdings durch die Akteure mit weniger Kontrolle getragen. Deshalb soll in Kapitel 5 erörtert werden, wann es zulässig ist, jemand anderem ein Risiko aufzuerlegen.

---

291 Hermansson, Hansson 2007:130

292 Im Folgenden wird erörtert, warum man aber die freie Wahl haben sollte und wie sich dies im Kontext von Datenökosystemen umsetzen ließe.

293 Luhmann 1990:161

294 Ebd.:149

295 Ebd.:153

3.3 Die Risiken der Akteure im Datenökosystem

Die Risiken entstehen auf Grund der epistemischen Ungewissheit über die Folgen des Umgangs mit Daten im Datenökosystem. Für alle vier Akteure ist es auf Grund möglicher Zweckentfremdung sowie undurchsichtiger Algorithmen schwierig abzusehen, welche Folgen das Sammeln und Verarbeiten der Daten haben kann. Die epistemische Asymmetrie verschärft die Ungewissheit ausgerechnet für die Akteure, die die Risiken tragen. Es ist aber erst die Kontrollungleichheit, die es den Akteuren mit weniger Wissen und Kontrolle (Datenquellen und Datennutzenden) verwehrt die Entscheidung zu treffen, ob sie die Risiken tragen wollen. Die Akteure mit mehr Wissen und Kontrolle (Datengenerierende und Datenverarbeitende) aber können die Entscheidung treffen, ob die Risiken eingegangen werden. Datengenerierende entscheiden, dass die Daten der Datenquellen gesammelt werden und wie viel Information sie über die Zwecke erhalten. Datenverarbeitende entscheiden, wie die Daten verarbeitet werden und wie die Ergebnisse für die Datennutzenden aufbereitet werden. Auf diese Weise legen Datengenerierende und Datenverarbeitende den Datenquellen und Datennutzenden Risiken auf.

Im Datenökosystem können durch die Kontrollungleichheit ebenfalls zwei unerwünschte Ereignisse mit einer gewissen Wahrscheinlichkeit eintreten. Sie sind den Risiken, die unter 3.1.1 aus einer rein epistemischen Perspektive betrachtet wurden, sehr ähnlich: Für Datenquellen besteht das Risiko, dass sie keine Kontrolle darüber haben, wer zu welchem Zweck auf ihre Daten zugreift. Für Datennutzende besteht das Risiko einem verzerrten Ergebnis zu glauben und somit zu falschen Überzeugungen zu gelangen, weil sie die Entstehung nicht überprüfen können. In beiden Fällen können die Akteure mit mehr Wissen und Kontrolle die Wahrscheinlichkeit, mit welcher das unerwünschte Ereignis eintritt, erhöhen oder verringern.

3.3.1 Risiko für Datenquellen

Ohne Datenquellen könnte es kein Datenökosystem geben – es gäbe keine Daten zum Sammeln und Verarbeiten. Trotzdem haben

Datenquellen kaum Kontrolle über den Datenfluss. Sie haben aber Interesse daran, zu kontrollieren zu welchen Zwecken ihre Daten genutzt werden und wer auf die Daten Zugriff hat. Dies ist ein Interesse an der eigenen informationellen Privatsphäre. Der Begriff der Privatsphäre wird in Kapitel 6 ausführlich definiert und seine Relevanz begründet. An dieser Stelle sei zunächst gezeigt, warum Datenquellen überhaupt um ihre Privatsphäre bangen müssen.

Während die Datenquellen ein Interesse daran haben, die Kontrolle über ihre Daten und damit ihre Privatsphäre zu behalten, haben Datengenerierende und Datenverarbeitende ein Interesse daran, möglichst viele Daten zu sammeln und zu verarbeiten. Das ist darin begründet, dass Datenverarbeitende sehr viele Daten benötigen, um ihre Algorithmen damit trainieren zu können – je mehr Daten, desto besser trainiert ist der Algorithmus. Diese Interessen geraten dadurch in Konflikt miteinander, dass es Zeit kostet, jede Person nach ihrer Zustimmung zu fragen und weil Personen ihre Daten vorenthalten und dadurch die Quantität der verfügbaren Daten verringern könnten.

3.3.2 Risiko für Datennutzende

Ohne die Datennutzenden müsste es kein Datenökosystem geben – wofür sollten Daten gesammelt und verarbeitet werden, wenn die Ergebnisse niemand nutzt? Trotzdem fehlt den Datennutzenden meist essentielles Wissen über die Entstehung der Ergebnisse, denen sie Glauben schenken. Es ist aber wichtig für sie zu erfahren, wie die Ergebnisse eines Datenökosystems zustande gekommen sind, um Fehler erkennen zu können. Dies ist ein Interesse an der transparenten Darstellung der Ergebnisse im Datenökosystem. Was genau Transparenz im Kontext eines Datenökosystems bedeutet und warum sie wichtig ist, wird in Kapitel 7 erörtert. An dieser Stelle sei zunächst gezeigt, warum die Transparenz im Datenökosystem nicht garantiert ist.

Hier besteht ein Konflikt zwischen dem Interesse an Transparenz der Datennutzenden und dem Interesse an wirtschaftlicher Rentabilität der Datenverarbeitenden. Laut dem Datenethiker Brent Mittelstadt und seinen Kolleg:innen ist dieser Interessenskonflikt geprägt von einer Informationsasymmetrie, da Datenverarbeitende

mehr Informationen darüber haben, wie die Ergebnisse zustande kommen.[296] Laut Mittelstadt et al. wird oft absichtlich geheim gehalten, wie genau der eingesetzte Algorithmus funktioniert, damit er nicht kopiert oder das Ergebnis des Algorithmus manipuliert wird und somit der Wettbewerbsvorteil verloren gehen kann (Gaming the System): „The commercial viability of data processors in many industries (e.g. credit reporting, high frequency trading) may be threatened by transparency."[297] Um marktfähig zu bleiben, seien manche Datenverarbeitende auf Geheimhaltung angewiesen – Transparenz würde ihr Geschäftsmodell gefährden.

Ein weiteres Problem ist, dass das Offenlegen der Funktionalität von verwendeten Algorithmen für die Datennutzenden, die selbst keine Programmiererfahrung haben, nicht zum Verständnis des Vorgangs beiträgt.[298] Trotzdem haben die Datennutzenden ein Interesse daran zu verstehen, wie die Ergebnisse, welche ihre Möglichkeiten oder ihre Entscheidungen beeinflussen, zustande kommen. Insbesondere weil durch dieses Verständnis Fehler durch verzerrte (biased) Daten erkannt werden können und Diskriminierung verhindert werden kann.

3.3.3 Zwischenfazit

Zusammengefasst kann man sagen, dass Datengenerierende durch das Sammeln und Speichern von personenbezogenen Daten das *Risiko* eingehen, dass digital gespeicherte Daten von Datenquellen zweckentfremdet werden und dass die Datenquellen die Kontrolle über ihre Daten verlieren – dadurch entsteht für Datenquellen eine *Gefahr*. Datenverarbeitende gehen durch das Analysieren der Daten mit Hilfe von KI-Systemen und das anschließende Bereitstellen von Ergebnissen das *Risiko* ein, dass die Ergebnisse für Datennutzende nicht mehr nachvollziehbar sind – bei essenziellen Themen, wie Wohngeld oder Berufswahl, kann dadurch eine *Gefahr* für die Datennutzenden entstehen.

296 Mittelstadt et al. 2016:6
297 Ebd.
298 Ebd.

Es ist wichtig zu betonen, dass die *Gefahr* im Datenökosystem dadurch entsteht, dass Datenquellen und Datennutzende bewusst durch Datengenerierende und Datenverarbeitende einem *Risiko* ausgesetzt werden. Denn Datengenerierende und Datenverarbeitende verursachen eine Situation, in welcher es möglich, aber nicht sicher ist, dass ein unerwünschtes Ereignis für Datenquellen und Datennutzenden eintritt – laut Hansson verursachen sie dementsprechend ein Risiko für andere.[299] Nun soll erörtert werden, wie diese Risikoauferlegung im Datenökosystem ethisch bewertet werden kann.

299 Hansson 2004:10

4 Ethische Theorien zur Bewertung von Datenökosystemen

In diesem Kapitel zeige ich, wie sich Datenökosysteme, als Risikoauferlegung zwischen Akteuren mit ungleicher Kontrolle, ethisch bewerten lassen und widme mich damit der Beantwortung von Forschungsfrage eins. Belliger und Krieger fordern eine Revolution der Ethik, die mit der Revolution der Technologie vergleichbar ist. Sie behaupten, dass die Ethik angesichts der digitalen Transformation selbst „digital" werden muss. Die Autor:innen verlangen insbesondere deshalb ein neues ethisches Denken, weil sie die modernen westlichen Traditionen in Frage stellen. Während sie die positiven Errungenschaften der westlichen Moderne (Rechtsstaatlichkeit, demokratische Prozesse, Freiheit, individuelle Rechte, Ziele wie Fairness, Gleichheit, Gerechtigkeit und persönlicher wie gesellschaftlicher Wohlstand) anerkennen und bewahren wollen, soll alles ihrer Ansicht nach Negative (Asymmetrien von Macht und Information, systemische Diskriminierung, kapitalistisches Wirtschaftssystem, Missachtung der Wahrheit in der politischen Kommunikation, territoriale Definition der Gesellschaft, Glaube an den Individualismus[300], Ökonomie der Aufmerksamkeit in den Medien) hinterfragt werden.[301]

Diese Probleme der westlichen Moderne seien durch das Beharren auf traditionellen Werten und ethischen Grundsätzen nicht gelöst, sondern sogar geschaffen oder gefördert worden: „This situation indicates that traditional values, norms, laws, and institutions are perhaps not the solution to the question of how society can best move into the future but could be distracting or even hinder-

300 Belliger und Krieger halten den Glauben an den Individualismus deshalb für negativ, weil er individuelle Interessen über das soziale Wohl stelle und alle äußeren Zwänge als Herausforderung für die individuelle Freiheit und Autonomie ablehne (Belliger, Krieger 2023:35f.).

301 Belliger, Krieger 2023:34–37

ing innovative attempts to solve these problems."[302] Weil also ethische Grundsätze (wie Menschenrechte, Menschenwürde, Autonomie, Privatsphäre, Schadensvermeidung, Fairness, Wohltätigkeit und Erklärbarkeit) die Probleme der westlichen Moderne nicht gelöst hätten, habe die Anwendung dieser Grundsätze auf neue Technologien keinen Sinn. Laut Belliger und Krieger ist es notwendig, über neue Werte, Normen und Formen der sozialen Ordnung nachzudenken, um konstruktiv in die Zukunft zu gehen.[303]

Ich argumentiere jedoch, dass neue Probleme und Fragestellungen nicht bedeuten, dass sich die Ethik an die neue Technologie anpassen muss. Ethik ist aus meiner Sicht nicht vorrangig zum Problemlösen da, sondern es ist die Aufgabe der Ethik, Probleme aufzudecken. Es bestehen genau da ethische Probleme, wo die ethischen Grundsätze missachtet werden. Außerdem muss man sich fragen, wie viel negativer die Situation wäre ohne die ethischen Grundsätze. Nur weil es gesellschaftliche Probleme gibt, während bestimmte Grundsätze herrschen, heißt das nicht, dass die Grundsätze die Ursache der Probleme sind. Achtung der Menschenwürde hat sicher nicht systemische Diskriminierung geschaffen oder gefördert – im Gegenteil: Systemische Diskriminierung widerspricht der Achtung der Menschenwürde. Genauso hat auch der Wert der Fairness nicht Asymmetrien von Macht und Information geschaffen oder gefördert, sondern wirkt ihnen direkt entgegen. Es gibt Probleme, und es gibt daneben ethische Grundsätze – Belliger und Krieger haben den von Ihnen vermuteten kausalen Zusammenhang nicht überzeugend dargelegt.

Sie haben aber noch weitere Einwände: Laut Belliger und Krieger können Werte aufgrund des historischen und sozialen Wandels nicht in ethischen Richtlinien festgeschrieben werden. Festgeschriebene Werte seien immer nur ein historischer Ausdruck der Werte einer bestimmten Gesellschaft zu einer bestimmten Zeit. Es sei jedoch ein Problem, dass sich die (digitale) Ethik heute an festgeschriebenen humanistischen Werten orientiere: „(…) ethics must move away from taking humanistic values for granted and attempting to apply them to new technologies and new situations."[304] An dieser Stelle

302 Belliger, Krieger 2023:37
303 Ebd.
304 Ebd.:41

wird deutlich, dass Belliger und Krieger Ethik mit Moral verwechseln. Denn Ethik sieht niemals Werte als selbstverständlich an – vielmehr ist es ihre Aufgabe, diese zu reflektieren. Ich stimme Belliger und Krieger zu, dass Digitalethik nicht den Zweck hat, Werte und Normen aus Richtlinien abzuleiten, und digitale Ethik ist ebenfalls keine philosophische Rechtfertigung für oktroyierte Richtlinien, Gesetze und Regierungsvorschriften.[305] Aber Ethik bedeutet auch nicht die Beschreibung der Normen eines Netzwerks.[306]

Ethik ist vielmehr die wissenschaftliche Untersuchung der Moral. Ihr Ziel ist es, die moralischen Überzeugungen der Menschen zu verstehen und zu rechtfertigen oder sie zu kritisieren und korrigieren. Es hat sich durchgesetzt die akademische Disziplin der Ethik in drei Bereiche zu unterteilen: Erstens die Metaethik (die Untersuchung dessen, was Moral ist), zweitens die Normative Ethik (das Studium moralischer Theorien und Konzepte) und drittens die Praktische Ethik (die Befassung mit ethischen Fragen, die sich im Rahmen bestimmter sozialer Praktiken stellen).[307] In dieser Arbeit werden der zweite und dritte Bereich eine Rolle spielen. In der Normativen Ethik werden durch das Studium ethischer Theorien Normen festgelegt, anhand derer wir Handeln bewerten können. Die Normen werden nicht aus Richtlinien abgeleitet und ergeben sich nicht aus der Beschreibung der zu bewertenden Gegenstände, sondern sie lassen sich mit Hilfe von wohldurchdachten und bewährten ethischen Theorien identifizieren. Die drei klassischen ethischen Theorien sind die Deontologie, der Konsequentialismus (z.B. Utilitarismus) und die Tugendethik. Im Folgenden soll erörtert werden, welche dieser Theorien am besten geeignet ist, um im Rahmen der Praktischen Ethik die Vorgänge in Datenökosystemen zu bewerten.

Die bestehenden ethischen Grundsätze, welche Belliger und Krieger kritisieren, kommen also nicht aus dem Nichts. Sie lassen sich ableiten aus wohldurchdachten ethischen Theorien, die einen gewissen Anspruch stellen an die Gesellschaft, in der wir miteinander leben wollen. So möchten Utilitarist:innen in einer Gesellschaft le-

305 Belliger, Krieger 2023:41, 52

306 Siehe Einleitung für die ausführliche Kritik dieses Alternativvorschlags von Belliger und Krieger.

307 Lindemann 2019:4f.

ben, in der die Lust maximiert und der Schmerz minimiert wird[308], Tugendethiker:innen wollen, dass sich die Menschen tugendhaft verhalten[309], und kantianische Deontolog:innen wollen eine Gesellschaft, in der alle Handlungsmaximen verallgemeinerbar sind[310] und so zum Beispiel niemand als bloßes Mittel gebraucht wird[311]. Neue Probleme und auch neue Technologien erfordern nicht, dass sich die Ethik ändert. Im Gegenteil, die neue Technologie muss den schon bestehenden ethischen Ansprüchen genügen. Es gibt nicht das eine ethische Prinzip, das alle Probleme löst. Aber die Anwendung der ethischen Theorien stellt Ansprüche an das menschliche Handeln unter der Annahme, sie wollen in einer Gesellschaft leben, in der die Lust maximiert wird oder sich alle tugendhaft verhalten beziehungsweise niemand als bloßes Mittel gebraucht wird[312].

Im Rahmen dieser Arbeit soll erörtert werden, wie Datenökosysteme ethisch gestaltet werden können. Was genau eine „ethische" Gestaltung ausmacht, hängt davon ab, welche Annahmen man macht über das menschliche Zusammenleben in einer Gesellschaft. Möchte man in einer Gesellschaft leben mit möglichst viel Lust und möglichst wenig Schmerz, dann liegt es nah, sich der utilitaristischen Nutzenkalkulation zu bedienen. Gleichzeitig ist es jedoch auch wichtig festzustellen, ob die ethische Theorie dazu in der Lage ist, den zu bewertenden Gegenstand hinreichend zu beleuchten. Da im Kontext dieser Arbeit Datenökosysteme bewertet werden sollen, muss die gesuchte ethische Theorie dazu in der Lage sein, Datenökosysteme samt ihrer Kontrollungleichheit und der Auferlegung von Risiken, zu beleuchten. Deshalb stellt sich nun die Frage, welche ethische Theorie sich für die Bewertung eines Risikos eignet.

Bisher bewertet die Moralphilosophie laut Hansson menschliches Verhalten in modellierten Szenarien mit vollständigem Wissen über die Situation. Die Entscheidungstheorie leite daraus Bewertungen für rationales Verhalten in realen, unsicheren Situationen ab. Hansson nennt dieses Vorgehen Arbeitsteilung zwischen Moralphiloso-

308 Bentham 2017 [1789]:7
309 Aristoteles 2006:Buch II, Kapitel 6, Abschnitt 1106b, Zeile 36 – Abschnitt 1107a, Zeile 6
310 Kant 2008:53 [421]
311 Ebd.:65 [429]
312 Ebd.:65 [429]

phie und Entscheidungstheorie.[313] Aber reale Unsicherheiten, wie zum Beispiel bei technologischen Entwicklungen oder dem Klimawandel, werfen nicht nur Entscheidungsprobleme auf, sondern auch ethische Probleme: „We need to have a moral view of the uncertainties of real life."[314] Hansson ist der Meinung, dass die strikte Unterteilung zwischen Moralphilosophie und Entscheidungstheorie keine ausreichenden Antworten liefert. Es muss auch möglich sein, die Unsicherheiten im echten Leben moralisch zu bewerten.

4.1 Erweiterung der Moraltheorien

Damit die gängigen ethischen Theorien wirksam auf Probleme angewendet werden können, die mit Risiko und Ungewissheit verbunden sind, müssen sie verallgemeinert werden auf ein Umfeld, das nicht deterministisch ist. Laut Hansson handelt es sich um „Kausale Verwässerung", wenn eine ethische Theorie so verallgemeinert wird, dass Wahrscheinlichkeiten eine Rolle spielen. Hansson formuliert das Problem der Kausalen Verwässerung (causal dilution problem) folgendermaßen: „Given that a certain moral theory prohibits a certain action because it has the property P, under what conditions should a generalized version of the same theory prohibit an action because it possibly, but not certainly, has the property P?"[315] Moraltheorien verbieten Handlungen auf Grund von bestimmten Eigenschaften – zum Beispiel gilt „Du sollst nicht töten" im Utilitarismus auf Grund der Konsequenzen oder bei Kant auf Grund der fehlenden Verallgemeinerbarkeit der Handlungsmaxime. Wenn nun aber unsicher ist, ob die zu bewertende Handlung, wie zum Beispiel Autofahren, eine solche verbotswürdige Eigenschaft hat, unter welchen Bedingungen sollte die Handlung verboten werden?

Haltaufderheide schreibt dazu: „Ob Risiken Grund sein sollen – ob sie Grund sein können – keine Kernkraftwerke zu bauen, dem Klimawandel entgegenzuwirken, nicht mehr Auto zu fahren oder keine Mobiltelefone zu benutzen, wissen wir nicht mit Sicherheit zu

313 Hansson 2003:291
314 Ebd.:292
315 Hansson 1999:541

sagen."[316] Denn die klassischen ethischen Theorien gehen bei ihrer Beurteilung einer Situation davon aus, dass man weiß, was passieren wird. Dann gibt es klare Auswahlmöglichkeiten zwischen möglichen konkreten Folgen. So ist offensichtlich, dass beim vielbesprochenen Trolley-Fallbeispiel, in welchem ein unaufhaltsamer Zug auf eine Gruppe von Menschen zurollt, der Zug entweder die Gruppe von Menschen sicher tötet oder auf ein anderes Gleis gelenkt wird, auf welchem ein Mensch sicher zu Tode kommt.[317] Dabei geht es den klassischen ethischen Theorien nie darum, wie hoch das Risiko für die Menschen ist. Es wird nicht gefragt, mit welcher Wahrscheinlichkeit die Menschen sterben oder nur verletzt werden – es wird vielmehr der schlimmste Fall (ihr Tod) angenommen. Es gibt aber auch Autor:innen, wie die normative Philosophin Frances Kamm[318], die sich mit Fragen des Risikos im Kontext des Trolley-Problems auseinandersetzten, doch das soll hier nicht betrachtet werden.

Denn in dieser Arbeit geht es nicht darum, unter welchen Bedingungen eine riskante Handlung verboten werden sollte, sondern darum, unter welchen Bedingungen anderen ein Risiko auferlegt werden darf. Es geht nicht darum herauszufinden, „(...) wie man sicher wissen kann, was man tun soll, wenn man nicht sicher wissen kann, was geschieht."[319], sondern darum herauszufinden, wann es zulässig ist, ein Risiko zu erzeugen, das andere tragen müssen. Laut Haltaufderheide ist das Problem der Kausalen Verwässerung der Ausgangspunkt für alle normativen Theorien des Risikos, und es erzeugt jedes Mal ein Dilemma. Seiner Meinung nach muss eine andere Frage gestellt werden und zwar, was Entscheidungssicherheit unter den Bedingungen der Unsicherheit heißen kann.[320]

Diese schwierige Frage, sowie die Frage danach, wie man sichere Urteile über Situationen unter Unsicherheit fällt, kann jedoch vernachlässigt werden, da in dieser Arbeit das Risiko selbst der Gegenstand ist, welcher ethisch bewertet werden soll. Es geht nicht darum, dass man bei der Bewertung eines Risikos (oder vielmehr seiner Unsicherheit) nicht weiß, was passieren wird. Sondern es geht

316 Haltaufderheide 2015:19
317 Foot 2002[1967]:23
318 Kamm 2020
319 Haltaufderheide 2015:18
320 Ebd.:24, 276

darum, dass bewusst eine Situation geschaffen wird, bei der es möglich ist, dass jemand zu Schaden kommt – es ist also sicher, dass ein Schadensrisiko besteht. Die Frage ist hier, ob dieses Schadensrisiko Akteuren, die ungleich weniger Kontrolle über eine Situation haben, auferlegt werden darf. Es geht also ganz konkret um die Gestaltung des menschlichen Miteinanders in dem besonderen Fall, dass Schaden möglich ist. Ein Risiko aufzuerlegen ist dabei natürlich nicht das Gleiche wie Schaden zu verursachen, da der antizipierte Schaden ja nur mit einer gewissen Wahrscheinlichkeit eintritt. Trotzdem bedarf auch die Tatsache, dass man einem Risiko durch andere ausgesetzt wird, einer ethischen Beurteilung.[321]

Wer ein Datenökosystem gestaltet, weiß mit Sicherheit, dass Datenquellen und Datennutzende dadurch Risiken tragen. Es handelt sich weder um eine Entscheidung unter Unsicherheit noch unter Risiko – es ist ganz eindeutig, dass Datenquellen und Datennutzende in Situationen gebracht werden, die zu unerwünschten Ereignissen führen können. Durch die Entwicklung von Datenökosystemen entstehen neue Risiken: Beim Einsatz von KI-Systemen entsteht das Risiko der fehlenden Transparenz, oder es werden bestehende Risiken verschärft, zum Beispiel durch die gesammelte Speicherung von Daten, die digital von überall zugänglich sind. Insbesondere weil das Risiko durch Datengenerierende und Datenverarbeitende geschaffen, aber von Datenquellen und Datennutzenden getragen wird, entsteht die Frage, ob diese Risikoschaffung, -verschärfung, und -auferlegung ethisch zulässig ist. Dabei geht es nicht darum, ob oder mit welcher Wahrscheinlichkeit die Risiken eintreten. Es geht nur darum, dass die Risiken den Datenquellen und Datennutzenden im Datenökosystem von Dritten auferlegt werden, und wie sich dies ethisch bewerten lässt. Zur Beantwortung dieser Frage wird im Folgenden eine ethische Theorie gesucht, welche die Zulässigkeit der Risiken im Datenökosystem bewerten kann. Zu diesem Zweck werden die drei klassischen ethischen Theorien des Utilitarismus, der Tugendethik und der Deontologie geprüft.

Die in Kapitel 2 und 3 erarbeitete Struktur eines Datenökosystems erhebt bestimmte Ansprüche an eine ethische Bewertung. Denn bewertet werden soll der Datenfluss zwischen vier Akteuren, von de-

321 Warum wird in Kapitel 5 ausführlich begründet.

nen zwei mehr Kontrolle haben über den Datenfluss als die anderen. Weil die Akteure mit mehr Kontrolle im Datenökosystem Risiken verursachen, welche die Akteure mit weniger Kontrolle tragen müssen, muss eine ethische Theorie, mit der Datenökosysteme bewertet werden sollen, ermöglichen

a) das asymmetrische Wissen sowie die ungleiche Kontrolle der Akteure im Datenökosystem zu betrachten und
b) das Auferlegen der Risiken zu bewerten.

4.2 Utilitarismus

Eine mögliche ethische Theorie für die ethische Bewertung von Datenökosystemen ist der Utilitarismus des Philosophen Jeremy Bentham. Er geht davon aus, dass Schmerz und Lust für das Handeln der Menschen ausschlaggebend sind: „Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. They alone point out what we *ought to* do and determine what we *shall* do (…).”[322] Laut Bentham bilden Schmerz und Lust den Maßstab für richtig und falsch und weisen die Menschen dementsprechend darauf hin, was sie tun sollten. Er ist der Meinung, dass die Menschen Schmerz und Lust unterworfen sind und entwickelt ein Prinzip, das diese Unterwerfung anerkennt und sie zur Grundlage eines Systems macht: „By ‘the principle of utility’ is meant the principle that approves or disapproves of every action according to the tendency it appears to have to increase or lessen—i.e. to promote or oppose—the happiness of the person or group whose interest is in question.”[323] Das Nützlichkeitsprinzip von Benthams Utilitarismus bewertet eine Handlung anhand des Schmerzes und der Lust, welche ihre Folgen verursachen. Mit Nutzen ist dabei die Eigenschaft einer Sache gemeint, entweder Lust zu erzeugen oder Schmerz zu verhindern. Es geht dem Utilitarismus also darum, das Glück der Personen, deren Interesse in Frage steht, zu fördern. Eine Sache fördert dann das Glück von Personen, wenn sie die Summe ihrer Lust erhöht und die Summe ihrer Schmerzen verringert. Eine

322 Bentham 2017 [1789]:6
323 Ebd.:6f.

Handlung stimmt also dann mit dem Nützlichkeitsprinzip überein, wenn ihre Tendenz, das Glück der betroffenen Personen zu erhöhen, größer ist als die Tendenz, es zu vermindern.[324]

Um bestimmen zu können, ob die Summe der Lust erhöht und die Summe der Schmerzen verringert wird, muss sowohl Lust als auch Schmerz ein Wert zugeschrieben werden. Für eine Person ist der Wert einer Freude oder eines Schmerzes größer oder kleiner, je nach Intensität, Dauer, Gewissheit oder Ungewissheit und Nähe oder Ferne. Zudem wird der Wert einer Freude oder eines Schmerzes größer oder kleiner, je nachdem, wie viele Personen davon betroffen sind.[325] Deshalb summiert man, um eine Handlung zu bewerten, die Werte aller durch sie erzeugten Lust und allen Schmerzes für jede Person, deren Interessen von der betrachteten Handlung betroffen zu sein scheinen: „The general tendency of an act is more or less pernicious according to the sum total of its consequences, i.e. according to the difference between the sum of its good consequences and the sum of its bad ones."[326] Für Bentham sind dabei die Konsequenzen der Handlung wichtig, welche aus Lust oder Schmerz bestehen oder Lust oder Schmerz erzeugen. Wenn man also mit Hilfe der ethischen Theorie des Utilitarismus eine Handlung (oder die Handlungen in einem Datenökosystem) bewertet, dann berechnet man mit Hilfe einer Art Kosten-Nutzen-Analyse, wie viel Lust oder Schmerz man bei der Ausführung dieser Handlung erwarten kann.[327]

Vallor ist der Meinung, dass der Utilitarismus als konsequentialistische Theorie im Kontext von technologischen Entwicklungen Schwierigkeiten hat, eine Handlungsweise zu empfehlen. Denn sie meint, jegliche Form von Konsequentialismus erfordere vollständige Informationen über die zu beurteilende Situation, um die Handlungsmöglichkeiten zu bewerten.[328] Beim Einsatz neuer Technologien sind jedoch nicht alle Konsequenzen vorhersehbar.[329] Da im Kontext von Datenökosystemen oft Informationen darüber fehlen,

324 Bentham 2017 [1789]:7
325 Ebd.:22
326 Ebd.:43
327 Keymolen, Taylor 2023:484
328 Vallor 2016:7f.
329 Keymolen, Taylor 2023:490

wer Zugang zu den Daten der Datenquellen erhält, und wie die Daten analysiert oder weitergegeben werden, ist diese Forderung hier nicht erfüllt. Weil wir nicht alle Konsequenzen kennen, die einzelne Handlungen in Datenökosystemen verursachen können, ist es laut Vallor mit der ethischen Theorie des Utilitarismus schwer zu ermitteln, wieviel Lust oder Schmerz die Folgen bestimmter Handlungen in einem Datenökosystem auslösen und ob sie dementsprechend moralisch richtig oder falsch sind. Die Undurchsichtigkeit der Konsequenzen des Einsatzes von Technologien mache es unkalkulierbar, welche Handlung das größte Glück für alle Beteiligten verspricht.[330]

Trotzdem ist Benthams Utilitarismus laut Keymolen und Taylor hilfreich bei der Bewertung von Handlungen in ungewissen Situationen. So könnte ein Utilitarist zwischen potenzieller Lust und potenziellem Schaden abwägen, wenn zum Beispiel im Kontext der Covid-19 Pandemie KI-Systeme zur Gesichtserkennung eingesetzt werden sollen, um die Personen zu identifizieren, die sich in der Nähe von infizierten Personen aufgehalten haben. Bei dem Einsatz einer solchen Technologie besteht einerseits das Risiko, dass Menschen die Kontrolle über ihre persönlichen Daten verlieren und ohne ihre Zustimmung identifiziert und kategorisiert werden. Andererseits könnte aber mit Hilfe dieser Technologie die Ausbreitung der Krankheit beschränkt werden, und es könnten womöglich Leben gerettet werden. Hier könnte man also die Werte der durch den Einsatz der Technologie erwartbaren Lust und des erwartbaren Schmerzes für jede betroffene Person summieren. Es wird angenommen, dass bei vielen Betroffenen große Lust entsteht durch das Verhindern von Krankheitsausbreitung und durch Lebensrettung. Gleichzeitig ist zu erwarten, dass einige Menschen unter der potenziellen Verletzung ihrer Privatsphäre leiden. Die Einführung dieser Gesichtserkennungstechnologie würde aber in diesem Fall mit dem Nützlichkeitsprinzip übereinstimmen, da sie tendenziell das Glück vieler betroffener Personen erhöht. und es nur für einige vermindert.[331]

Ein Utilitarist wäre also auch dazu in der Lage, das Auferlegen von Risiken ethisch zu bewerten. Dies ist nach dem Nützlichkeitsprinzip dann verwerflich, wenn es mehr Schmerz als Lust erzeugt. Für den

330 Vallor 2016:7f.
331 Keymolen, Taylor 2023:484f.

Utilitaristen wäre es also von Bedeutung, wie hoch das Schadensrisiko und die Eintrittswahrscheinlichkeit sind. Zur Berechnung von Lust und Schmerz, die durch das Auferlegen eines Risikos eintreten könnten, gibt es die moderne Risikoanalyse. Diese basiert auf einem utilitaristischen Ansatz, der durch die Erwartungsnutzentheorie ergänzt wird. Letztere bemisst ein Risiko, indem der Schweregrad der negativen Ereignisse, auf die sich das Risiko bezieht, mit deren Eintrittswahrscheinlichkeit gewichtet wird.[332] Ein bestehendes Risiko wird gewichtet durch das Produkt aus Wahrscheinlichkeit und Schweregrad des Schadens – so errechnet sich der statistische Erwartungswert.[333] Es erfolgt also eine Quantifizierung des potenziellen Schadens zum Beispiel als Geldbetrag, der zur Reparatur oder als Schadensersatz gezahlt werden müsste, oder auch als Anzahl von Menschenleben, die gefährdet sein könnten. Diesen quantifizierten Schaden multipliziert man dann mit der vermuteten Eintrittswahrscheinlichkeit. Ebenso lässt sich das für den potenziellen Nutzen anwenden, der durch die Auferlegung der Risiken erwartet wird.[334]

Hermansson und Hansson üben Kritik an diesem Verständnis von Risiko: „The ethics of risk-taking and risk imposition concerns problems of agency and interpersonal relationships that cannot be adequately expressed in a framework that operates exclusively with the probabilities and severities of outcomes."[335] Der utilitaristische Ansatz mit Ergänzung durch die Erwartungsnutzentheorie könne Probleme des Handelns und der zwischenmenschlichen Beziehungen in Entscheidungen unter Risiko nicht angemessen adressieren. Denn die Kosten-Nutzen-Analyse berücksichtigt nicht, wer diese Kosten trägt und wer den Nutzen einfährt.[336] Nutzen (Lust) und Kosten (Schmerz) unterschiedlicher Akteure werden addiert, ohne dass die Tatsache berücksichtigt wird, dass sie an verschiedene Personen gebunden sind. Daher können Kosten (Schmerz), die eine Person betreffen, durch einen ausreichend großen Vorteil für eine andere Person gerechtfertigt werden. So kann der Nutzen (Lust) für eine Person die Risikoexposition einer anderen Person (Schmerz)

332 Hermansson, Hansson 2007:129
333 Hansson 2004:10; Wolff 2021:5
334 Ropohl 2016:11f.
335 Hermansson, Hansson 2007:130
336 Wolff 2021:11

überwiegen.[337] In unsicheren Entscheidungssituationen sind aber laut Hansson nicht nur die möglichen Konsequenzen moralisch relevant, sondern auch, wer die Konsequenzen trägt: „A disadvantage for one person is not necessarily outweighed by an advantage for another person, even if the advantage is greater than the disadvantage."[338] Um die ethische Perspektive in die Risikoanalyse einzubeziehen, müssen wir also analysieren, wer wem welche Risiken auferlegt und mit welcher Absicht.[339]

Laut Ropohl werden zudem bei der utilitaristischen Nutzenkalkulation moralisch relevante Schäden durch die Verknüpfung mit der Eintrittswahrscheinlichkeit relativiert. Denn eine Handlung, die mit Sicherheit den Tod einer Person herbeiführt, ist zu unterlassen. Wenn aber die gleiche Handlung zulässig wird, weil die Todesfolge nur mit geringer Wahrscheinlichkeit eintritt, „(...) dann tut man so, als wenn die niedrigere Eintrittswahrscheinlichkeit die Handlung von der moralischen Verwerflichkeit ihrer möglichen Folge reinigen würde."[340] Insbesondere aber, weil es schwer ist, die Eintrittswahrscheinlichkeit von Handlungsfolgen richtig einzuschätzen und es jederzeit möglich ist, dass (unerwartet) Menschen zu Schaden kommen[341], sollte die geringe Eintrittswahrscheinlichkeit eines existenziellen Risikos laut Ropohl kein Grund sein für die Rechtfertigung der Auferlegung eines solchen Risikos.[342]

Zusammenfassend kann ich sagen, dass die ethische Theorie des Utilitarismus nicht die Richtige ist, um Datenökosysteme zu bewerten. Für einen Utilitaristen ist die oben postulierte Bedingung (b) das Auferlegen von Risiken dann zulässig, wenn es mehr Lust als Schmerz erzeugt und somit insgesamt das Glück aller betroffenen Personen erhöht. Allerdings könnte das Auferlegen von Risiken dann auch zulässig sein, wenn ausschließlich diejenigen davon profitieren, die das Risiko auferlegen, solange ihre Lust insgesamt größer ist als der Schmerz der Risikoexponierten. Denn das Problem in der Risikoanalyse und im Utilitarismus allgemein ist, dass Vertei-

337 Hansson 2004:25
338 Hansson 2005:10
339 Hansson 1999:542; Hansson 2003:302; Hermansson, Hansson 2007:130
340 Ropohl 2016:18
341 Luhmann 1990:158; Ropohl 2016:18
342 Ropohl 2016:18

lungsfragen außer Acht gelassen werden.[343] Es ist aber essenziell zu betrachten, wer im Datenökosystem das Risiko auferlegt, und wer es trägt, um zum einen (a) das asymmetrische Wissen sowie die ungleiche Kontrolle der Akteure im Datenökosystem betrachten zu können und zum anderen (b) das Auferlegen der Risiken bewerten zu können. Außerdem braucht ein Utilitarist zur Bewertung des Risikoauferlegens Informationen darüber, wie viel Schmerz und wieviel Lust dadurch erzeugt wird. Für ihn sind relevant die Höhe des Schadensrisikos, die Eintrittswahrscheinlichkeit sowie ihre Multiplikation zum Erwartungsnutzen. Dabei misst der, durch die Erwartungsnutzentheorie ergänzte, Utilitarismus der Eintrittswahrscheinlichkeit eines Schadens einen so hohen Stellenwert zu, dass unwahrscheinliche aber moralisch relevante Schäden relativiert werden.[344] Somit ist auch durch den Fokus auf die Konsequenzen der Risiken anstatt auf die Beziehung zwischen den Akteuren, die das Auferlegen ermöglicht, das Bewerten des Auferlegens der Risiken (b) erschwert.

4.3 Tugendethik

Die ethische Theorie der Tugendethik bewertet die Zulässigkeit von Handlungen nicht anhand ihrer Konsequenzen, sondern anhand der Charaktereigenschaften der Handelnden. Hier soll die Tugendethik des Philosophen Aristoteles betrachtet werden, da er als einer der Gründerväter der modernen Tugendethik gilt.[345] Laut Aristoteles streben die Menschen nach einem guten Leben (eudaimonia). Das gute Leben besteht dabei in der guten und schönen Ausführung aller Handlungen mit Vernunft, was ermöglicht wird durch die Aneignung von Tugenden. Die Tugend des Menschen definiert Aristoteles als „(...) diejenige Disposition, durch die er ein guter Mensch wird und durch die er seine Funktion gut erfüllt."[346] Tugenden sind, genau wie Laster, nach Aristoteles erworbene, stabile Eigenschaften. Aristoteles unterscheidet zwischen intellektuellen (oder rationalen)

343 Hansson 2004:26
344 Ropohl 2016:18
345 Keymolen, Taylor 2023:486; Beier et al. 2024:50
346 Aristoteles 2006:Buch II, Kapitel 5, Abschnitt 1106a, Zeile 23–24

und ethischen Tugenden: Intellektuelle Tugenden, wie Weisheit und Klugheit, werden durch Belehrung erworben und vervollkommnen das menschliche Denkvermögen. Ethische Tugenden, wie Freigebigkeit und Besonnenheit, werden hingegen durch das Nachahmen von Vorbildern erworben und sie vervollkommnen den sinnlich strebenden Teil des Menschen.[347] Ethisch tugendhaft ist man dann, wenn man wissend, vorsätzlich und aus einer festen und unveränderlichen Haltung heraus gut handelt.[348] Dabei liegt eine ethische Tugend immer in der Mitte zwischen zwei Lastern – zwischen Mangel und Übermaß. Während lasterhafte Handlungen entweder hinter dem Gesollten zurückbleiben oder darüber hinausgehen, finden tugendhafte Handlungen genau die Mitte und wählen das Gesollte.[349] Die Tugend „Mut" zum Beispiel wird definiert als genau die richtige Menge an Furcht, die man in gefährlichen Situationen empfinden sollte – bei zu viel Furcht handelt man feige und bei zu wenig Furcht tollkühn. Wer aber genau die richtige Menge an Furcht empfindet, erkennt, was er zu tun hat, und handelt mutig.[350]

Vallor vertritt die Auffassung, dass die Tugendethik die richtige ethische Theorie zur Bewertung von technologischen Entwicklungen ist, weil sie einen Rahmen bietet für die Pluralität ethischer Theorien und sie die Kultivierung von technomoralischen Gewohnheiten und Tugenden erleichtert.[351] Ihrer Meinung nach kann die Tugendethik als eine ethische Strategie eingesetzt werden, bei der der moralische Charakter hilft, eine unsichere Situation zu bewältigen.[352] Nur die tugendhaften Personen mit praktischer Weisheit können richtig auf die einzigartigen und wechselnden Anforderungen jeder konkreten moralischen Situation abgestimmt handeln.[353] Vallor identifiziert zwölf Tugenden, die einen ethisch korrekten Umgang mit neuen Technologien erleichtern sollen: Ehrlichkeit, Selbstkontrolle, Bescheidenheit, Gerechtigkeit, Tapferkeit, Empathie, Fürsor-

347 Aristoteles 2006:Buch I, Kapitel 13, Abschnitt 1102b.1103a
348 Ebd.:Buch II, Kapitel 3, Abschnitt 1105a, Zeile 29–34
349 Ebd.:Buch II, Kapitel 6, Abschnitt 1106b, Zeile 36 – Abschnitt 1107a, Zeile 6
350 Beier et al. 2024:51
351 Vallor 2016:9f.
352 Ebd.:10
353 Ebd.:26

ge, Kooperationsbereitschaft, Flexibilität, Augenmaß, Großmut und Weisheit.[354]

Die ethische Theorie der Tugendethik kann laut Keymolen und Taylor tatsächlich hilfreich sein bei der Bewertung von Handlungen in ungewissen Situationen. Im Jahr 2018 haben Google-Mitarbeitende dagegen protestiert, dass ihr Unternehmen bei der Verbesserung der Analyse von Drohnenvideos des US-amerikanischen Verteidigungsministeriums half. Sie hatten die Befürchtung, dass diese Technologie eingesetzt werden könnte, um menschliche Ziele für Luftangriffe auszuwählen. Da sich nach unternehmensintern geäußerten Bedenken nichts geändert hatte, haben die Mitarbeitenden eine Petition unterzeichnet und manche von ihnen haben aus Protest gekündigt. Das Projekt wurde schließlich eingestellt. Hier waren die Mitarbeitenden weder feige noch tollkühn, sondern sie haben mutig gehandelt angesichts der Befürchtung, dass die Technologie, die ihr Unternehmen entwickelte, zum Töten verwendet werden könnte.[355]

Die Tugendethik kann durch das Beurteilen des moralischen Charakters auch Situationen bewerten, in denen Risiken bestehen, wie zum Beispiel die eventuelle Nutzung einer Technologie für schädliche Zwecke. Ein Tugendethiker würde das Auferlegen von Risiken dann für verwerflich halten, wenn es aus lasterhaftem Verhalten geschieht. Denn aus der Sicht der Tugendethik sollte jede involvierte Person Tugenden ausbilden und dementsprechend Haltungen einnehmen, die allen zugutekommen sollen. Dies soll den Individuen ermöglichen, angemessen auf ethische Herausforderungen zu reagieren und die Wahrscheinlichkeit erhöhen, dass bei der Entwicklung und dem Einsatz von Technologien ethisch gehandelt wird.[356] Wo tugendhaft gehandelt wird, kann ein vertrauensvoller Umgang miteinander entstehen, aber das Vorkommen von tugendhaftem Verhalten allein garantiert kein Vertrauen. Denn es ist oft so, dass nicht jede Person alle relevanten Tugenden ausbilden kann oder will.[357]

Hier stößt der tugendethische Ansatz an seine Grenzen, da er keinen Unterschied macht zwischen den Akteuren mit mehr und mit weniger Kontrolle. Im Datenökosystem aber besteht ja ein Ungleich-

354 Vallor 2016:120; Übersetzung angelehnt an Beier et al. 2024:55
355 Keymolen, Taylor 2023:486f.
356 Hagendorff 2020:113; Beier et al. 2024:58
357 Beier et al. 2024:56

gewicht des Wissens und der Kontrolle. Da die Datengenerierung für die Datenquellen und die Datenverarbeitung für die Datennutzenden meist schwer nachzuvollziehen und kaum zu kontrollieren sind, müssen die Akteure mit weniger Kontrolle (Datenquellen und Datennutzende) sich darauf verlassen, dass die Akteure mit mehr Kontrolle (Datengenerierende und Datenverarbeitenden) tugendhaft handeln. Ein tugendhafter Charakter mag helfen, mit technologischen Unsicherheiten umzugehen, aber er schützt nicht vor Gefahren, die von nicht tugendhaften Personen verursacht werden. Man kann mit Hilfe der Tugendethik Datengenerierende und Datenverarbeitende nicht dazu verpflichten, sich tugendhaft zu verhalten. Und selbst wenn alle Akteure des Datenökosystems einen tugendhaften Charakter haben, könnte eine außenstehende, nicht tugendhafte Person Zugang zu den Daten erhalten und sie zweckentfremden.

Die Relevanz des asymmetrischen Wissens sowie der ungleichen Kontrolle zeigt sich auch an dem soeben angeführten Beispiel: Angesichts des Google-Projektes mit Drohnenvideos des US-amerikanischen Verteidigungsministeriums mussten erst viele tugendhafte Mitarbeitende das Wissen über das Projekt erlangen und öffentlich Bedenken äußern, bevor das Projekt eingestellt wurde.[358] Es wäre aber schon vor dem Beginn des Projektes möglich gewesen, Bedenken bezüglich der Nutzung der Technologie zu haben beziehungsweise die unternehmensintern geäußerten Bedenken ernst zu nehmen – hier liegt das eigentliche Problem. Denn die Führungsebene bei Google scheint sich bewusst für das bedenkliche Projekt und gegen das Reagieren auf Bedenken entschieden zu haben. Diejenigen, die über die Aufnahme und Einstellung des Projektes entscheiden, haben keine Fürsorge gezeigt für die potenziellen Opfer ihrer Technologie, und sie haben durch das Ignorieren der firmeninternen Bedenken keine Empathie oder Kooperationsbereitschaft ihren Mitarbeitenden gegenüber gezeigt. Es ist zwar lobenswert, dass die Mitarbeitenden – als Akteure mit weniger Kontrolle – tugendhaft gehandelt haben, aber es ist bedeutsamer, ob die Akteure mit mehr Kontrolle tugendhaft handeln. Wenn sie erst mit Petitionen und Kündigungen dazu gebracht werden müssen, ist die Tugendethik nicht ausreichend. Denn man kann zwar an den tugendhaften Charakter der Akteure mit mehr Wissen und Kontrolle appellieren,

358 Keymolen, Taylor 2023:486f.

aber die Theorie bildet nicht ab, dass es auf die Akteure mit mehr Wissen und Kontrolle ankommt, einen tugendhaften Charakter auszubilden, weil es ihre Entscheidungen sind, die anderen Risiken auferlegen können.

Mein Schluss ist, dass die ethische Theorie der Tugendethik nicht die Richtige ist, um Datenökosysteme zu bewerten. Sie erfüllt zwar Bedingung (b), das heißt, sie kann das Auferlegen der Risiken bewerten anhand des Charakters derjenigen, welche die Risiken auferlegen, aber sie beachtet nicht (a) das asymmetrische Wissen sowie die ungleiche Kontrolle der Akteure im Datenökosystemen, weil die Beziehung zwischen den Datenquellen und Datengenerierenden sowie Datennutzenden und Datenverarbeitenden nicht betrachtet wird.

4.4 Kants Deontologie

Aus Kants Sicht sind wir Menschen Teil zweier unterschiedlicher Welten: der Sinnenwelt und der intelligiblen Welt. Als Teil der Sinnenwelt sind wir Naturwesen, welche von Naturgesetzen bestimmt werden. Das heißt, wir folgen unseren Neigungen und handeln zum Beispiel, weil wir Durst oder Hunger haben – wir handeln auf Grund von Begierde. Wenn wir das tun, reagieren wir auf ein Bedürfnis, das wir nicht frei gewählt haben. Wir Menschen sind als Teil der Sinnenwelt in unserem Handeln fremdbestimmt, da wir Zwecke verfolgen, deren Urheber wir nicht sind.[359] Bei solchem Handeln orientieren wir uns an hypothetischen Imperativen, welche sich unserer instrumentellen Vernunft bedienen und nur von für den jeweils spezifischen Fall gültig sind. Wir handeln, um ein bestimmtes Ziel zu erreichen oder ein Bedürfnis zu befriedigen – die Handlung wird zum Mittel.[360]

Als Teil der intelligiblen Welt sind wir Menschen Vernunftwesen, die zur Autonomie fähig sind.[361] Wir sind dazu in der Lage, aus Pflicht zu handeln, das heißt, weil die Handlung richtig ist und nicht, weil sie nützlich oder üblich ist. Das Motiv der Pflicht verleiht

359 Kant 1956
360 Kant 2008:44 [414]
361 Kant 1956

unserem Handeln moralischen Wert und ermöglicht Sittlichkeit. Als Teil der intelligiblen Welt sind wir Menschen in unserem Handeln selbstbestimmt, da wir nach Gesetzen handeln, die wir uns selber geben. Wir handeln frei, wenn wir den Zweck als solchen wählen, um seiner selbst willen. Um selbstbestimmt und frei zu sein, müssen wir auf Grund eines Kategorischen Imperativs handeln. Kategorische Imperative sind Gesetze, die wir uns selber geben mit Hilfe unserer Vernunft. Erst wenn die Vernunft unseren Willen bestimmt, können wir unabhängig von unseren Neigungen entscheiden. Wenn wir uns unserer Vernunft bedienen, kommen wir Menschen laut Kant alle zu dem gleichen Schluss – dem Kategorischen Imperativ. Ein Kategorischer Imperativ zeichnet sich dadurch aus, dass die an ihm orientierte Handlung an sich gut ist – sie ist sogar notwendig als Prinzip eines vernunftgemäßen Willens.[362]

Kants erste Formulierung des Kategorischen Imperativs fordert, dass jede menschliche Handlung so betrachtet werden sollte, als würde sie zu einem Gesetz für alle werden: „Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde."[363] Dies ist ein Test der Universalisierung, um zu sehen, ob die Maxime von allen gedacht und gewollt werden kann. Wenn eine Handlungsmaxime weder gedacht noch gewollt werden kann, ist es eine vollkommene Pflicht, diese Handlung zu vermeiden. So dürfen zum Beispiel keine Versprechen gemacht werden unter dem Vorsatz, diese nicht einzuhalten. Denn würde die Handlungsmaxime verallgemeinert, verlöre jedes Versprechen seine Bedeutung – das ist weder denkbar, noch kann es gewollt werden.[364] Eine Handlungsmaxime, die zwar denkbar ist, aber nicht widerspruchsfrei gewollt werden kann, etabliert eine unvollkommene Pflicht. So sollten Menschen, die anderen Menschen in Not helfen könnten, ihre Hilfe nicht vorenthalten. Die Verallgemeinerung einer solchen Handlungsmaxime wäre zwar denkbar, aber man kann es nicht wollen, da man selber im Notfall keine Hilfe bekäme.[365]

Kants dritte Formel des Kategorischen Imperativs bezieht sich konkreter auf den Umgang mit anderen Menschen: „Handle so,

362 Kant 2008:35f. [408], 46f. [416]; Sandel 2009:144–191
363 Ebd.:53 [421]
364 Ebd.:55f. [422]
365 Ebd.:57f. [423f.]

daß du die Menschheit sowohl in deiner Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst."[366] Jeder Mensch hat also die Pflicht, andere Menschen nicht bloß als Mittel zu gebrauchen, sondern immer ihren Selbstzweck anzuerkennen. Denn „(...) vernünftige Wesen haben Würde."[367] Aus diesem Grund schulden wir allen vernünftigen Wesen die Pflicht zur Achtung. Laut dem Moralphilosophen Richard Hare ist Kant deshalb ein Deontologe, weil er der Pflicht in seiner Darstellung des moralischen Denkens einen vorrangigen Platz einräumt.[368]

Limitiert ist der kantianische deontologische Ansatz zum Beispiel durch Kants Bedingung der Vernunftfähigkeit. Laut Kant haben wir Menschen Würde, weil wir vernünftige Wesen und Urheber der Gesetze sind, nach denen wir handeln.[369] Zwar ist nicht jeder Mensch im gleichen Maße in der Lage, sich seiner oder ihrer Vernunft zu bedienen, aber ich möchte trotzdem sagen, dass alle Menschen gleichermaßen eine Würde haben. In dem Anwendungsfall dieser Arbeit, der Digitalisierung der Verwaltung, lässt sich aber die unbedingte Voraussetzung der Vernunft verteidigen. Diejenigen, bei denen hier Vernunft und damit eine Fähigkeit zur Sittlichkeit vorausgesetzt wird, sind die Akteure, welche bewusst Daten generieren und verarbeiten. Es ist die Pflicht der Datengenerierenden und Datenverarbeitenden zum moralischen Handeln, welche im Folgenden betrachtet wird. Bei diesen Akteuren setze ich eine Vernunftfähigkeit voraus.

In der Philosophie geht eine Denkrichtung davon aus, dass in Kants Deontologie die Konsequenzen der Handlungen irrelevant seien. So schreibt zum Beispiel der Philosoph Hans Jonas über Kants Kategorischen Imperativ: „In der Tat, *reale* Folgen sind überhaupt nicht ins Auge gefaßt (...)."[370] Hier wird behauptet, dass die Konsequenzen einer Handlung für Kants Deontologie unwichtig sind. Kent selbst macht tatsächlich einen Unterschied zwischen Handlungen aus Pflicht und Handlungen aus Besorgnis um die nachteiligen

366 Kant 2008:65 [429]
367 Sandel 2009:168
368 Hare 1993:2
369 Kant 2008
370 Jonas 2020[1979]:37

Folgen – Erstere sind geboten, Letztere hingegen verboten, wenn sie dem Kategorischen Imperativ widersprechen.[371] Allerdings muss ja nicht jede Handlung aus Pflicht der Besorgnis um die nachteiligen Folgen widersprechen.

So begründet Hare nachvollziehbar, dass auch aus deontologischer Sicht manche Handlungsfolgen relevant sind – und zwar sind genau die Folgen relevant, welche die moralischen Grundsätze verbieten oder herbeiführen wollen.[372] So sei zum Beispiel Lügen deshalb falsch, weil es dazu führt, dass jemand anderes getäuscht wird und sich eine falsche Überzeugung bildet. Die dritte Formel des Kategorischen Imperativs verbietet aber, dass andere Menschen in dieser Weise als bloßes Mittel gebraucht werden. Diese beabsichtigte Konsequenz, andere getäuscht zu haben (und dadurch bloß als Mittel zu gebrauchen), sei das, was Lügen falsch macht. Ich soll also die Wahrheit sagen, auch wenn das die Konsequenz hat, dass ich dadurch benachteiligt werde – denn mein Nachteil sei nach dem Kategorischen Imperativ keine moralisch relevante Konsequenz. Auf die gleiche Weise sei das Töten eines Menschen deshalb falsch, weil es den Tod dieses Menschen zur Folge hat. Der Grundsatz des Kategorischen Imperativs verbietet die Handlungsfolge des Todes, weil die zugrundeliegende Handlungsmaxime weder gedacht noch gewollt werden kann und somit eine vollkommene Pflicht entsteht, diese Handlung zu vermeiden. Entsprechend dieses Grundsatzes ist es moralisch relevant, wenn die Folge einer Handlung der Tod einer Person ist.[373] Selbst wenn in Kants Deontologie die aus den Kategorischen Imperativen abgeleiteten Pflichten ausschlaggebend sind für die Bewertung von Handlungen, sind die Folgen der Handlungen dennoch nicht irrelevant. In der Deontologie sind nur die Folgen relevant, die vom Kategorischen Imperativ geboten oder verboten werden. Kants Deontologie beantwortet die Frage danach, was grundsätzlich gelten muss, damit die Gestaltung eines Datenökosystems prinzipiell ethisch gut ist – dazu gehört auch, dass moralisch relevante, negative Konsequenzen vermieden werden.

Laut Vallor hat Kants Deontologie jedoch Schwierigkeiten zu entscheiden, ob die Vorgänge in Datenökosystemen moralisch richtig

371 Kant 2008:28 [403]
372 Hare 1993:15
373 Ebd.

sind, weil die praktischen Ungewissheiten es unmöglich machen zu wissen, ob man eine Zukunft mit Datenökosystemen wollen kann: „(…) you need to know what will be *done* with the [data], *who* might control them, and *how* they would be assessed or shared."[374] Ohne diese Details zu kennen, sei es schwierig, zukünftige technologische Entwicklungen ethisch zu bewerten. Ich argumentiere jedoch, dass Kants deontologischer Ansatz lediglich Kenntnis von den Beziehungen zwischen den Akteuren braucht, um eine ethische Bewertung vornehmen zu können. Wer personenbezogene Daten sammelt und verarbeitet, hat die Pflicht, die Datenquellen nicht bloß als Mittel zu gebrauchen. Wenn aber die Daten der Datenquellen verwendet werden, ohne dass diese davon wissen oder Kontrolle darüber haben, dann werden sie als bloßes Mittel gebraucht. Wer die Ergebnisse von verarbeiteten Daten bereitstellt, hat die Pflicht, die Datennutzenden nicht bloß als Mittel zu benutzen. Wenn aber Datennutzende Ergebnissen ausgesetzt werden, ohne dass sie wissen wie diese zustande gekommen sind oder Kontrolle darüber haben, dann werden sie als bloßes Mittel gebraucht.[375]

Um ethische Bewertungen auch in Situationen unter Risiko ausführen zu können, gibt es einige Ideen für die Erweiterung von deontologischen Theorien. Zum Beispiel könnte man postulieren, ein Verbot, ein bestimmtes Ergebnis herbeizuführen, impliziert ein Verbot, die Wahrscheinlichkeit dieses Ergebnisses zu erhöhen (auch wenn die Erhöhung sehr gering ist).[376] Leider ist eine solch strikte Ausweitung von Verboten laut Hansson gesellschaftlich unhaltbar: „Such a strict interpretation would make human society impossible."[377] Denn das hieße unter anderem, niemand dürfte mehr Auto fahren, weil dadurch das Risiko steigt, dass Menschen zu Schaden kommen. Deontologische Theorien verbieten manche Handlungen, die allgemeine Pflichten verletzen, aber nicht alle solcher Handlungen.[378]

374 Vallor 2016:7

375 Wann genau jemand als bloßes Mittel gebraucht wird, wird unter 5.2.3 ausführlich erörtert.

376 Hansson 2003:298

377 Ebd.:298

378 Hansson 1999:542; Hansson 2003:298

Eine Idee zur Integration des Faktors Unsicherheit ist das Verbinden einer jeden Pflicht mit einer Wahrscheinlichkeitsgrenze. Unterhalb dieser Grenze ist das Risiko erlaubt und oberhalb der Grenze ist es verboten. Allerdings ist es laut Hansson sehr schwierig, diese Grenze zu ziehen: „(…) probability limits would draw the line between acceptable and unacceptable probabilities of harm with no regard to the benefits involved."[379] Eine pauschale Risikogrenze wäre nicht in der Lage, die Vorteile der jeweiligen risikobehafteten Handlung einzubeziehen.

Eine weitere Idee ist das Einführen einer Regel, die den Pflichten Gewichtungen zuweist, wobei die Gewichte mit abnehmender Risikowahrscheinlichkeit kleiner werden.[380] Das heißt, wenn eine Handlung nur mit geringer Wahrscheinlichkeit einen Schaden verursacht, ist die Pflicht, die Handlung zu unterlassen, nur gering gewichtet. Laut Hansson führt diese Gewichtung allerdings zu weit weg von den deontologischen Theorien. Stattdessen entsteht eine Form des erweiterten Utilitarismus, da die Verletzung eines Rechts oder die Übertretung eines Verbots einen negativen Wert hat und in einem Nutzenkalkül abgewogen wird. „It is not clear that this construction is at all a non-empty moral theory."[381] Hansson befürchtet, dass eine Erweiterung der deontologischen Theorien durch Gewichtung zu einer leeren Moraltheorie führt.

Ich verstehe jedoch die Nutzung der deontologischen Theorien für einen Risikokontext etwas anders. Aus meiner Sicht ist es nicht notwendig, sich in den Konflikt zu begeben, dass deontologische Verbote oder Gebote unter Risiko nur noch mit einer gewissen Wahrscheinlichkeit gelten könnten. Stattdessen sollen deontologische Pflichten dabei helfen, das sichere Auferlegen von Risiken (b) zu bewerten. Ein kantianischer Deontologe würde das Auferlegen von Risiken dann für verwerflich halten, wenn die Risikoexponierten dadurch als bloßes Mittel gebraucht werden oder wenn sich die zugrundeliegende Handlungsmaxime anderweitig nicht verallgemeinern ließe. Hier ist nicht relevant, ob das Risiko eintritt oder wie hoch der Schaden wäre, und auch die Intention derjenigen, die das Risiko auferlegen, ist zweitrangig. Vielmehr ist die Beziehung zwischen den Akteuren relevant.

379 Hansson 2003:299
380 Ebd.
381 Ebd.

Das zeigen Keymolen und Taylor an folgendem Beispiel: Im Jahr 2020 wurde in Den Haag ein von Regierungsbehörden eingerichtetes System zur Erstellung von Risikoprofilen für Sozialbetrug gerichtlich verboten. Das System hat Personen anhand von Daten über ihr Einkommen und Wohneigentum bis hin zu Daten über ihren Wasser- und Energieverbrauch bewertet und die Wahrscheinlichkeit, mit der sie Sozialbetrug begehen, errechnet. Dabei wussten die Bürger:innen nicht, was mit ihren Daten geschieht, dass sie überhaupt anhand von ihren Daten bewertet werden und wie das System zu seinen Einstufungen kommt. Zum einen war dieses System gesetzeswidrig in den Niederlanden, zum anderen lässt sich das Verbot eines solchen Systemen auch mit Hilfe von Kants Deontologie begründen. Denn wenn ein solches System die Daten von Bürger:innen analysiert und sie einer Bewertung unterzieht, ohne dass die Bürger:innen dies wissen, nachvollziehen oder verweigern können, werden sie bloß als Mittel gebraucht.[382]

In diesem Beispiel wurden den Bürger:innen der Niederlande Risiken auferlegt – sie wurden durch den Einsatz des Systems in eine Situation gebracht, in der es möglich, aber nicht sicher ist, dass ein unerwünschtes Ereignis eintreten wird.[383] Dies ist deshalb signifikant, weil man behaupten könnte, die Wahrscheinlichkeit, dass unerwünschte Ergebnisse eintreten, wurde durch das System ohne Not erhöht, da die Suche nach Sozialbetrüger:innen auf diesem Wege für den geregelten Ablauf des alltäglichen Lebens nicht notwendig erscheint.[384] Kants Deontologie ermöglicht es, die Aufmerksamkeit auf die Beziehung zwischen den Akteuren zu richten. Denn seine ethische Theorie zeigt auf, wer wen als bloßes Mittel gebraucht.

Zusammenfassend komme ich zu der Einschätzung, dass die ethische Theorie der kantianischen Deontologie die Richtige ist, um Datenökosysteme zu bewerten. Denn sie ermöglicht es zu betrachten, wer im Datenökosystem das Risiko auferlegt, und wer es trägt. Außerdem sind für den deontologischen Ansatz die Höhe des Schadensrisikos und seine Eintrittswahrscheinlichkeit nicht relevant. Stattdessen ermöglicht der deontologische Blick die Konzentration auf die Beziehung zwischen den Akteuren im Datenökosystem. So ist

382 Keymolen, Taylor 2023:485
383 Hansson 2004:10
384 Ropohl 2016:17

es möglich zum einen (a) das asymmetrische Wissen sowie die ungleiche Kontrolle der Akteure im Datenökosystem zu betrachten und zum anderen (b) das Auferlegen der Risiken zu bewerten. Wenn Akteure mit mehr Wissen und Kontrolle dazu neigen, ihr Wissen und ihre Kontrolle auszunutzen und die Akteure mit weniger Wissen und Kontrolle bloß als Mittel zu gebrauchen, macht Kants Deontologie darauf aufmerksam. Prinzipiell gilt für alle die gleiche Regel, doch es wird erkennbar, dass es für Akteure mit mehr Wissen und Kontrolle einfacher ist, diese zu brechen, und es womöglich sogar Anreize dafür gibt.

Kants Deontologie erfüllt als einzige der drei klassischen ethischen Theorien beide von mir aufgestellten Kriterien, die zur Bewertung von Datenökosystemen notwendig sind. Deshalb soll sie im Folgenden als Grundlage dienen für die ethische Bewertung und Anforderungen an die Gestaltung von Datenökosystemen. Zuerst wird aber genauer betrachtet, was das Auferlegen von Risiken für die Akteure bedeutet.

5 Zulässigkeit der Auferlegung von Risiken

Um das Argumentationsziel dieser Arbeit zu erreichen, und zu zeigen, wie der Umgang mit digitalen Daten in Datenökosystemen ethisch gestaltet werden kann, muss beantwortet werden, wann die Struktur von Datenökosystemen ethisch zulässig ist. Da die Struktur von Datenökosystemen von der durch die Kontrollungleichheit ermöglichten Risikoauferlegung bestimmt wird, wird in diesem Kapitel erörtert, unter welchen Bedingungen es zulässig ist, anderen ein Risiko aufzuerlegen. Zuerst wird die ethische Relevanz von Risikoauferlegung etabliert, dann wird eine Pflicht zur Vermeidung der Risikoauferlegung begründet und anschließend wird eine Ausnahme für diese Pflicht vorgeschlagen. In Ermangelung von ausreichenden Regeln für diese Ausnahme, schlage ich Kants dritte Formel des Kategorischen Imperativs als Regel vor und diskutiere was genau das bedeutet. Für mögliche Umsetzungsschwierigkeiten in der Praxis werden zum Schluss Lösungsangebote gemacht.

Das Verursachen eines Risikos im Datenökosystem ist nichts Außergewöhnliches. Im Alltag werden in vielen Situationen Risiken verursacht, die für andere zur Gefahr werden können. Dem Medien- und Technologiephilosophen Mark Coeckelbergh zufolge sind alle menschlichen Aktivitäten und Praktiken mit Risiken verbunden: „Driving a car and traffic, like all human activities and practices, involves risk."[385] Jede Person, die im Straßenverkehr Auto fährt, erzeugt Risiken, die für andere Verkehrsteilnehmende Gefahren darstellen können. Trotzdem nehmen wir regelmäßig am Straßenverkehr teil und verursachen Risiken oder setzen uns Gefahren aus. Das Verursachen von Risiken lässt sich nicht vermeiden, da die Zukunft ungewiss ist und jederzeit selbstverschuldete unerwünschte Ereignisse eintreten können. Das menschliche Leben kann nicht risikofrei sein, und das sollte es auch nicht. Den politischen Philosophen Jonathan Wolff und Avner De-Shalit zu folge kann die Of-

385 Coeckelbergh 2016:752

fenheit für Risiken sogar ein Vorteil sein. Vermeidbare Risiken, die man bewusst eingeht, indem man zum Beispiel eine neue Beziehung eingeht, sind Teil dessen, was ein menschliches Leben ausmacht. Risiken werden regelmäßig bewusst eingegangen, zum Beispiel um ein glückliches Leben führen zu können oder in der Hoffnung, dass man dadurch einen Gewinn erzielt.[386]

Man kann dann durch Risiken beeinträchtigt werden, wenn der riskierte Schaden eintritt oder aber die Risiken unfreiwillig eingegangen werden. Unfreiwillig bedeutet hier nicht, dass Gewalt angewendet wurde, sondern dass es keine (gute) Alternative gab: „(…) very often people are disadvantaged because they are exposed to risks which they would not have taken had they had the option (…)."[387] In solchen Fällen können die Betroffenen wählen, ob sie ein Risiko eingehen oder vermeiden wollen, aber wenn sie es nicht eingehen, sind sie in bestimmten Fällen mit einem größeren Risiko oder der Gewissheit eines Schadens konfrontiert.[388] So sind Wohngeldantragstellende zum Beispiel gezwungen, der Weitergabe ihrer Daten zur Erstellung der Wohngeldstatistik zuzustimmen. Dies kann deshalb ein Risiko darstellen, weil es dadurch wahrscheinlicher wird, dass Unbefugte Zugriff auf die Daten bekommen und man als Teil der Gruppe der von Armut Bedrohten abgewertet und dementsprechend diskriminiert wird. Es gibt aber keine gute Alternative zum Eingehen dieses Risikos, weil man ohne die Zustimmung keinen Wohngeldantrag stellen kann. Wer aber auf die staatliche Unterstützung durch Wohngeld angewiesen ist, ist in diesem Sinne gezwungen das Risiko einzugehen.

Der Jurist und Philosoph John Oberdiek ist der Meinung, dass das Auferlegen von Risiko durch solchen Zwang moralisch signifikant wird: „The idea, at bottom, is that imposing risk on others is morally significant in these terms because risk impositions amount to claims of authority over others' lives, and more specifically, over the range of options that constitutes their autonomy."[389] Wenn man anderen Personen ein Risiko auferlegt, fordert man laut Oberdiek Autorität über das Leben dieser Personen, da man ihre Handlungsoptionen

386 Wolff, De-Shalit 2007:66
387 Ebd.
388 Ebd.:67
389 Oberdiek 2017:90

und damit ihre Autonomie beschränkt. Autonomie lässt sich laut der Ethikerin Beate Rössler folgendermaßen definieren: „(...) autonom ist eine Person dann, wenn sie sich die Frage stellen kann, welche Person sie sein will, wie sie leben will, und wenn sie dann so leben kann."[390] Damit die Autonomie gelingt, muss die Person fähig sein, ihre Wünsche zu reflektieren und sie anzunehmen, abzulehnen oder anzupassen[391] und bei der Ausbildung dieser Wünsche darf keine direkt intendierte Manipulation oder Repression im Spiel sein.[392] Denn welche Ziele eine Person verfolgen kann, ist abhängig von ihrem Umfeld.[393] So kann das Auferlegen eines Risikos als Repression verstanden werden, wenn es nicht mit den Wünschen der Person übereinstimmt. Denn die Risikoexponierten werden hierbei fremdbestimmt und können in diesem Kontext nicht mehr so leben, wie sie wollen.

Durch eine solche Risikoauferlegung entsteht Schaden – kein materieller Schaden, sondern immaterieller Schaden an der Autonomie der betroffenen Personen.[394] Diese repressiven externen Bedingungen können dazu führen, dass sich die betroffene Person über ihre Wünsche, Ziele und Möglichkeiten falsche Überzeugungen bildet und es so zu einem Mangel an Autonomie kommt.[395] Zu einem solchen Mangel sollte es nicht kommen, da die Autonomie eine notwendige Bedingung für ein gelungenes Leben ist.[396]

Darüber hinaus äußert sich solch immaterieller Schaden durch das Auferlegen eines Risikos zum Beispiel durch Stress und Ängste bei den Risikoexponierten.[397] Denn allein die Tatsache, dass man in der Gegenwart unsicher ist, ob in der Zukunft ein Schaden eintritt, ist laut Luhmann „(...) in vielen Hinsichten ein bereits gegenwärtiger Schaden."[398] Auch der politische Philosoph Jonathan Wolff stimmt dem zu: „(...) there is a cost to living with risk or uncertainty, even

390 Rössler 2001:39
391 Ebd.:103
392 Ebd.:119
393 Ebd.:108
394 Oberdiek 2017:87
395 Rössler 2001:118
396 Ebd.:96,126
397 Wolff, De-Shalit 2007:68
398 Luhmann 1990:159

if nothing bad ever happens."[399] Die Technologieethikerin Sabine Roeser hat, passend dazu, im Jahr 2014 das Paradox der unerträglichen Ungewissheit (Unbearable Uncertainty Paradox) identifiziert. Dieses entsteht, wenn Menschen die Ungewissheit über den Ausgang einer Situation so sehr ablehnen, dass sie lieber einen Schaden erleiden würden, als den ungewissen Zustand auszuhalten. Auch wenn man sich bewusst ist, dass die Präferenz des Schadens irrational ist, wird der psychologische Schmerz, den die Ungewissheit auslöst, als schlimmer empfunden, als die Erfahrung des Schadens.[400]

Das Auferlegen von Risiken ist also nicht nur auf Grund der riskierten Schäden ethisch relevant, sondern allein schon wegen dem möglichen Mangel an Autonomie sowie der aufgezwungenen Unsicherheit, welche selbst als Schaden empfunden werden kann.

5.1 Bedingungen für die Auferlegung von Risiken

Da sich das Auferlegen von Risiken nicht insgesamt vermeiden lässt, muss ein guter Umgang damit gefunden werden. Oberdiek geht davon aus, dass es in bestimmten Kontexten zulässig ist, anderen ein Risiko aufzuerlegen. Aber diese Auferlegung des Risikos beeinträchtigt die Autonomie einer Person, wenn sie über ihren Umgang mit der Möglichkeit von Schädigungen (Risiko) nicht selbst bestimmen kann. Da eine solche fremdbestimmte Risikoauferlegung also einen immateriellen Schaden darstellen kann, muss sie gerechtfertigt werden.[401] Hannson formuliert die Frage, die sich hier stellt, in Form des Ausnahmeproblems (The exemption problem): "It is a prima facie moral right not to be exposed to risk of negative impact, such as damage to one's health or one's property, through the actions of others. What are the conditions under which this right is overridden, so that someone is allowed to expose other persons to risk?"[402] Unter welchen Bedingungen dürfen Individuen einem Risiko negativer Auswirkungen ausgesetzt werden? Ich gehe hier davon aus, dass bei der Risikoauferlegung nicht die Wünsche der Risikoexponierten

399 Wolff 2021:12
400 Roeser 2014:640, 648
401 Oberdiek 2017:87
402 Hansson 2003:303

berücksichtigt werden.[403] Das Recht, keinem Risiko ausgesetzt zu sein, wird durch eine solche bedingte Ausnahme allerdings nicht eingeschränkt, sondern nur außer Kraft gesetzt. Deshalb hat dieses Recht immer noch Einfluss und kann die Begründung der Ausnahme schwächen. Außerdem kann das Recht, keinem Risiko ausgesetzt zu sein, laut Hansson dazu führen, dass bei Ausnahmen zusätzliche Pflichten entstehen: „(...) such as obligations to compensate, to inform, and to reduce the risk whenever possible."[404] Hier stellt sich jedoch die Frage, wie sich das Recht, keinem Risiko ausgesetzt zu sein, begründet.

In der philosophischen Debatte herrscht Uneinigkeit darüber, ob es solch ein moralisches Recht überhaupt geben kann. Der praktische Philosoph Georg Mohr zum Beispiel vertritt die Meinung, dass es keine moralischen Rechte gibt: „Es gibt kein Recht, zumal kein verlierbares Recht, auf Wahrheit bzw. Wahrhaftigkeit seitens des Hörers einer Behauptung. Es gibt nur die Pflicht seitens des Sprechers wahrhaftig zu sprechen."[405] Moral und Recht seien zwei verschiedene Kategorien. Im Recht ginge es nicht um Moral, und in der Moral ginge es nicht um Rechte. Die Verknüpfung der beiden Bereiche zu moralischen Rechten sei ein Kategorienfehler.[406] Mohr argumentiert zudem, dass es moralische Rechte nicht gibt, weil wir einerseits keine klare Definition von ihnen haben, und es andererseits in jedem Anwendungsbereich bessere Alternativen gibt.[407] „Sie sind keine moralischen Rechte, sondern aus unbedingten moralischen Pflichten resultierende Primärrechte, die als Legitimitäts-Forderungen an Rechtsordnungen gerichtet sind."[408] Die Alternative zu moralischen Rechten sind demnach moralische Pflichten.

Mohr ist davon überzeugt, dass es nur moralische Pflichten gibt, einander auf eine bestimmte Art und Weise zu behandeln. Jegliches Recht, auf eine bestimmte Art und Weise behandelt zu werden, sei aus der moralischen Pflicht abgeleitet. Es handle sich dabei aber

403 Anders ist das natürlich, wenn die Risikoexponierten gleichzeitig auch die Entscheidungstragenden sind und sich auf Grund ihrer eigenen Wünsche Risiken aussetzen.
404 Hansson 2003:303
405 Mohr 2010:68
406 Ebd.:71
407 Ebd.:78
408 Ebd.:77

nicht um moralische Rechte, sondern um juridische Rechte, mit welchen die Menschen zur Einhaltung ihrer moralischen Pflichten gezwungen werden können.[409] Moralische Pflichten sind den praktischen Philosoph:innen Corinna Mieth und Christoph Bambauer zufolge „(...) strikt verbindliche Sollensforderungen, die sich das praktische Subjekt selbst auferlegt."[410] Laut Mieth und Bambauer lassen sich aus vollkommenen Pflichten Rechte ableiten, die (juridisch) geltend gemacht werden können, aber die Erfüllung von unvollkommenen Pflichten kann nicht eingefordert werden, da die Einforderung von Hilfspflichten zur Überforderung führen würde.[411] Um Konflikte mit dem Rechtsbegriff zu vermeiden, wird im Folgenden angenommen, dass es moralische Pflichten gibt und sich jegliche Rechte aus diesen ableiten lassen.

Die hier relevante Frage ist, ob es Pflichten der Unterlassung oder der besonderen Rechtfertigung der Auferlegung von Risiken gibt. Das von Hansson postulierte Recht, keinem Risiko auf Schaden ausgesetzt zu werden, ließe sich also nur ableiten aus einer vollkommenen moralischen Pflicht, andere Menschen keinem Schadensrisiko auszusetzen. Wer handelt, muss gegenüber den Betroffenen sicherstellen, keine Pflichtverletzung zu begehen. Wenn es sich um einen mit Sicherheit eintretenden Schaden handeln würde, der aus der Sicht der Betroffenen nicht zu rechtfertigen ist, kann mit Hilfe des Kategorischen Imperativs gezeigt werden, dass die zugrunde liegende Maxime unmoralisch ist.[412] Bei der Auferlegung eines Risikos tritt der Schaden allerdings nur mit einer gewissen Wahrscheinlichkeit ein. Deshalb ist es wichtig, sich nicht auf den riskierten Schaden als mögliche Folge zu konzentrieren, sondern auf den Schaden, der durch das gezwungene Auferlegen des Risikos mit Sicherheit entsteht.

Bei der ethischen Bewertung des Aktes der Auferlegung eines Risikos, kann die dritte Formel von Kants Kategorischem Imperativ hilfreich sein: „Handle so, daß du die Menschheit sowohl in deiner

409 Mohr 2010:71, 77

410 Mieth, Bambauer 2016:3

411 Ebd.; Es ist eine vollkommene Pflicht, eine Handlung zu vermeiden, wenn die Handlungsmaxime weder gedacht noch gewollt werden kann. Eine Handlungsmaxime, die zwar denkbar ist, aber nicht widerspruchsfrei gewollt werden kann, etabliert eine unvollkommene Pflicht (Kant 2008:55f. [422])

412 Kant 2008:53 [421]

Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst."[413] Es kann weder gewollt noch gedacht werden, dass man andere Menschen bloß als Mittel gebraucht. Aus diesem Anspruch an das Zusammenleben entspringt diese moralische Forderung an das Verhalten: Wenn du in einer Gesellschaft leben willst, in der die Menschen einander nicht als bloßes Mittel benutzen, dann erkennst du damit die moralische Pflicht an, Handlungen zu vermeiden, durch welche andere Menschen bloß als Mittel gebraucht werden. Wenn aber Entscheidungstragende Akteuren mit weniger Kontrolle ein Risiko auferlegen, und die nun Risikoexponierten eigentlich das Risiko gerne vermeiden würden, es dazu aber keine gute Alternative gibt, dann werden die Risikoexponierten von den Entscheidungstragenden bloß als Mittel gebraucht. Denn den Risikoexponierten wird die Möglichkeit genommen, sich gegen oder frei für das Risiko zu entscheiden. So lässt sich die These von Hanssons Ausnahmeproblem begründen:

(1) Wir haben eine Pflicht das eigene Handeln gegenüber anderen so zu gestalten, dass die anderen niemals bloß als Mittel, sondern immer auch als Selbstzweck behandelt werden (Kants dritte Formel des Kategorischen Imperativs).

(2) Wenn andere einem Risiko auf Schaden ausgesetzt werden und ihnen dabei die Möglichkeit zur Kontrolle über das Risiko verwehrt wird, werden sie nicht als Selbstzweck behandelt.

(3) Wenn man von anderen gegen seinen Willen und ohne zureichende Aufklärung einem Risiko auf Schaden ausgesetzt wird, wird einem die Möglichkeit zur Kontrolle über das Risiko verwehrt.

→ Es ist eines jeden Pflicht, durch das eigene Handeln andere keinem Risiko auf Schaden gegen seinen Willen und ohne zureichende Aufklärung auszusetzen.

Aus dieser Pflicht der Unterlassung von Risikoauferlegung lässt sich eine Pflicht der besonderen Rechtfertigung der Auferlegung von Risiken ableiten. Denn es stellt sich weiterhin die Frage, unter welchen Bedingungen eine Person trotzdem einem Risiko ausgesetzt werden

413 Kant 2008:65 [429]

darf. Unter solchen Bedingungen würde eine Ausnahme gemacht und die Pflicht, andere keinem Risiko auszusetzen, würde außer Kraft gesetzt.[414] Dieses Vorgehen lässt sich alltäglich beobachten, doch unter welchen Umständen halten wir es für akzeptabel? Laut Hansson würde die Risikoanalyse folgendermaßen auf die Frage nach der Ausnahme für die Risikoauferlegung antworten: „A risk imposition is acceptable if the total benefits that it gives rise to outweigh the total risks, measured as the probability-weighted disutility of outcomes."[415] Hansson wendet jedoch ein, dass die Pflicht, Personen keine Risiken aufzubürden, nicht allein dadurch untergraben wird, dass andere Akteure dadurch begünstigt würden – auch nicht, wenn die positive Auswirkungen für die Begünstigten größer sind als die negativen Auswirkungen für die Risikoexponierten.[416] Hier zeigt sich erneut das Problem, dass bei der utilitaristischen Risikoanalyse jede risikoexponierte Person als Träger von Nutzen und Schaden behandelt wird, deren Wert gleich bleibt, auch wenn sie von jemand anderem getragen würden. In dieser Arbeit wird jedoch Wert darauf gelegt, dass jede risikoexponierte Person als autonomes Individuum behandelt wird, welches nach seinen eigenen Wünschen leben kann.

Hansson stellt ein anderes Prinzip auf, welches die Bedingungen beschreiben soll für eine Ausnahme zu der Pflicht, niemanden einem Risiko auszusetzen: „Exposure of a person to a risk is acceptable if and only if this exposure is part of an equitable social system of risk-taking that works to her advantage."[417] Eine Person darf dann und nur dann einem Risiko ausgesetzt werden, wenn dieses Aussetzen Teil eines gerechten sozialen Systems der Risikobereitschaft ist, von dem sie selbst profitiert. Ein solches gerechtes soziales System der Risikobereitschaft enthält typischerweise Regeln, die bestimmte Arten von risikobehafteten Aktivitäten zulassen, ohne dass im Einzelfall eine Kosten-Nutzen-Abwägung erforderlich ist. Diese Regeln sollen für alle von Vorteil sein.[418] Dadurch wird zumindest anerkannt, dass jede risikoexponierte Person ein autonomes Individuum

414 Hansson 2003:303
415 Hansson 2004:32
416 Hansson 2018:1827
417 Hansson 2003:305
418 Ebd.:306

ist, welches nach seinen eigenen Wünschen leben möchte. Wenn diese Person einem Risiko ausgesetzt wird, müssen ausreichende Gründe für die fremdbestimmte Akzeptanz dieses Risikos angegeben werden.[419]

Hansson beschreibt ein solches gerechtes soziales System der Risikobereitschaft als „Risikoaustausch", bei dem sich die Menschen gegenseitig Risiken aussetzen und das für alle Beteiligten von Vorteil ist.[420] Solange also alle Beteiligten trotz eines Risikos gleichermaßen von der Maßnahme begünstigt sind, ist das Risiko, und somit die Gestaltung der Technologie, ethisch zulässig. Das heißt, ein Mensch darf Auto fahren im Rahmen eines sozialen Systems mit Regeln, auch wenn dadurch das Risiko eines Schadens für alle anderen Verkehrsteilnehmenden entsteht. Das soziale System sorgt dafür, dass jede autofahrende Person einen Führerschein besitzt und somit gelernt hat, welche Regeln im Straßenverkehr gelten. Zudem darf niemand betrunken Auto fahren, weil das die anderen Verkehrsteilnehmenden einem unzumutbar erhöhten Risiko aussetzen würde.[421] Das Risiko, das durch Autofahrende entsteht, wird nur akzeptiert, weil jede Person gleichermaßen davon profitieren könnte, indem sie selbst Auto fährt oder sich fahren lässt. Dabei kann das risikoexponierte Individuum seinen Vorteil allerdings anders einschätzen, als eine unabhängige Einschätzung es vorschlagen würde.[422]

Ich argumentiere, dass es für ein gerechtes soziales System nicht reicht, wenn die Risikoauferlegung für alle Beteiligten von Vorteil ist. Denn selbst wenn alle Beteiligten vom Risiko Begünstigte sind, sind bei einer Risikoauferlegung die Entscheidungstragenden nicht identisch mit den Risikoexponierten. So zum Beispiel beim Wohngeldantrag, bei welchem die Wohngeldbehörden oder im weitesten Sinne der Staat (Entscheidungstragende) den Antragstellenden (Risikoexponierte) das Risiko auferlegen, dass ihre Daten potenziell zweckentfremdet werden können bzw. der zur Antragbearbeitung genutzte Algorithmus ihren Antrag fälschlicherweise ablehnen könnte. Dabei sind zwar alle Akteure durch das Risiko begünstigt, aber die Risi-

419 Hansson 2003:306
420 Hansson 2018:1827
421 Ropohl 2016:17
422 Hansson 2003:305f.

koexponierten können das Risiko nicht frei wählen.[423] Den Risikoexponierten wird von den Entscheidungstragenden ein Risiko auferlegt, das sie nicht kontrollieren können – das widerspricht einem „gerechten sozialen System".

Im Alltag setzen wir Menschen uns gegenseitig Risiken aus, aber dies wird nicht einfach so hingenommen. Die Akzeptanz ist geknüpft an die Bedingung, dass die Risikoauferlegung innerhalb eines gerechten sozialen Systems der Risikobereitschaft stattfindet, von dem auch die Risikoexponierten profitieren. Im Folgenden wird eine weitere Regel aufgestellt, welche ein gerechtes soziales System der Risikobereitschaft ermöglicht – auch wenn die Risikoexponierten nicht gleichzeitig die Rolle der Entscheidungstragenden einnehmen.

5.2 Eine Regel für ein gerechtes soziales System der Risikobereitschaft

Hansson bleibt sehr vage, wenn es darum geht, das gerechte soziale System der Risikobereitschaft zu definieren. Wenn wir aber wissen wollen, wie das Auferlegen von Risiken im Datenökosystem ethisch gestaltet werden kann, dann müssen wir wissen, wann ein soziales System der Risikobereitschaft gerecht ist. Deshalb soll in diesem Kapitel Kants dritte Formel des Kategorischen Imperativs als Regel für ein solches System eingeführt werden. Dadurch wird im Folgenden eine deontologische Risikoethik begründet.

Laut Oberdiek sind Autofahrer deshalb verpflichtet, nicht gefährdend zu fahren, weil die potenziellen Unfallopfer ein Interesse daran haben: „Drivers are duty-bound not to drive dangerously, and the reason why drivers are so bound is due to the interests of would-be victims of a crash."[424] Er schlussfolgert, dass sich diese Pflicht durch ein Recht (keinem Risiko ausgesetzt zu werden) der potenziellen Opfer begründen müsse. Aus meiner Sicht aber entsteht die Pflicht, nicht gefährdend Auto zu fahren, nicht durch die Interessen der möglichen Opfer – sie entsteht vielmehr durch ein Interesse aller

423 Hermansson, Hansson 2007:142
424 Oberdiek 2017:100

Menschen in einer Gesellschaft zu leben, in welcher sich alle gegenseitig als Selbstzweck und nicht bloß als Mittel benutzen.[425]

Deshalb könnte die dritte Formel des Kategorischen Imperativs von Kant eine der grundlegenden Regeln sein, die laut Hansson ein gerechtes soziales System der Risikobereitschaft ausmachen: „The social systems of risk-taking (…) typically include rules, such as those permitting automobile traffic, that allow certain types of risk-creating activities without requiring appraisals of costs and benefits in the individual cases."[426] Kants dritte Formel des Kategorischen Imperativs würde nur eine bestimmte Art von risikobehafteten Aktivitäten zulassen – nämlich nur solche Aktivitäten, die andere Menschen als Selbstzweck und nicht als bloßes Mittel benutzen.[427] Mit Hilfe dieser Regeln wären im Einzelfall keine Kosten-Nutzen-Abwägung erforderlich, solange alle Menschen immer auch als Selbstzweck behandelt werden. Zudem wäre diese Regel für alle Menschen gleichermaßen von Vorteil.

Hansson spricht von Regeln, die beispielsweise den Autoverkehr ermöglichen. Darunter stellt man sich erstmal konkretere Regeln vor, wie zum Beispiel den notwendigen Erwerb eines Führerscheins oder das Verbot des alkoholisierten Fahrens. Aber jede dieser praktischen Regeln lässt sich zurückführen auf ein Gebot der Rücksichtnahme auf andere Verkehrsteilnehmende – und damit auf die dritte Formel des Kategorischen Imperativs von Kant. Diese Formulierung des Kategorischen Imperativs eignet sich in besonderer Weise für eine Begründung des Gebots der Rücksichtnahme, sowohl im Verkehr als auch im Umgang mit digitalen Daten. Denn, sie konzentriert sich auf die Beziehung zwischen den Akteuren und stellt eine Regel auf für die gegenseitige Behandlung: Sie verbietet es einander als bloßes Mittel zu gebrauchen und sie gebietet es einander als Selbstzweck zu behandeln. Die dritte Formel des Kategorischen Imperativs macht also darauf aufmerksam, dass man bei jeder Handlung darauf achten sollte, wie man seine Mitmenschen behandelt – man sollte Rücksicht auf die von der Handlung betroffenen Menschen nehmen. Da diese Rücksichtnahme gerecht und sozial wäre, ist die dritte Formel des Kategorischen Imperativs besonders geeignet als Regel für ein gerechtes soziales System der Risikobereitschaft.

425 Kant 2008:65 [429]
426 Hansson 2003:306
427 Kant 2008:65 [429]

5.2.1 Behandlung als Selbstzweck

Um die dritte Formel des Kategorischen Imperativs von Kant als Regel eines gerechten sozialen Systems der Risikobereitschaft zu implementieren, muss festgestellt werden, was es heißt, so zu handeln „(...) daß du die Menschheit sowohl in deiner Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst."[428] Widmen wir uns zunächst dem Gebot des Brauchens von anderen als Selbstzweck – die Philosophiehistorikerin Pauline Kleingeld, die sich vor allem mit den Texten von Kant befasst, versteht dies folgendermaßen: „Die Formulierung ‚als Zweck gebrauchen' ist wohl am besten im Sinne von „*qua* Zweck" oder „gemäß seinem Status als Zweck gebrauchen" zu verstehen, d.h. als Hinweis darauf, dass man einen Menschen so gebrauchen soll, wie es seinem moralischen Status als Selbstzweck entspricht."[429] Man kann laut Kant jemanden als Selbstzweck gebrauchen, indem man seine bzw. ihre Zwecke respektiert und anerkennt: „Denn das Subjekt, welches Zweck an sich selbst ist, dessen Zwecke müssen, wenn jene Vorstellung bei mir *alle* Wirkung tun soll, auch, so viel möglich, *meine* Zwecke sein."[430] Das heißt, ein Subjekt als Zweck an sich anzuerkennen bedeutet, seine Zwecke als Zwecke vom selben Rang wie die eigenen zu betrachten, sodass die Zwecke des anderen als ebenso verbindlich gelten wie die eigenen. Wenn ich mein Gegenüber benutze, sollte ich laut Hare seine Zwecke als meine Zwecke behandeln: „(…) in order to fulfil this version of the Categorical Imperative, I have to treat other people's ends (i.e. what they will for its own sake) as *my* ends."[431] Man kann womöglich nicht jeden Zweck eines jeden Gegenübers kennen, aber man kann sich zumindest über Zwecke mit einem fundamentalen Rang Gedanken machen. So kann man zum Beispiel davon ausgehen, dass jeder Mensch den Wunsch hat weiterzuleben, und man sollte diesen Zweck der anderen in seine eigene Handlungsentscheidung miteinbeziehen. Ebenso kann man sich vorstellen, dass das Gegenüber meines Handelns möglichst kein Risiko auferlegt bekommen möchte, oder, wenn es selbst auch von der Risikoauferlegung profitiert, ein Interesse an einer möglichst

428 Kant 2008: 65 [429]
429 Kleingeld 2014:117
430 Kant 2008:67 [430]; Hervorhebungen im Zitat gesperrt.
431 Hare 1993:4

geringen Eintrittswahrscheinlichkeit und Schadenshöhe des Risikos hat.

Gegen dieses Gebot verstößt man laut Kant zum Beispiel dann, wenn man anderen nicht hilft: „Nun würde zwar die Menschheit bestehen können, wenn niemand zu des andern Glückseligkeit was beitrüge, dabei aber ihr nichts vorsetzlich entzöge; allein es ist dieses doch nur eine negative und nicht positive Übereinstimmung zur *Menschheit als Zweck an sich selbst*, wenn jedermann auch nicht die Zwecke anderer, so viel an ihm ist, zu befördern trachtete."[432] Damit eine Handlung also positiv übereinstimmt mit der Menschheit als Zweck an sich selbst, müsste sie die Zwecke anderer befördern. Durch seine Handlungen soll man also nicht nur der Menschheit als Zweck an sich nicht im Weg stehen, sondern man soll diese Selbstzweckhaftigkeit auch befördern. Geboten ist also nicht nur (negativ) die Unterlassung der Missachtung der Zwecke anderer, sprich die Unterlassung der Schädigung von Risikoexponierten durch die Auferlegung von Risiken, sondern es ist auch (positiv) geboten, Datenökosysteme so zu gestalten, dass die Zwecke aller Menschen befördert werden, die sie als Subjekte qua Zweck an sich selbst haben.

5.2.2 Benutzung als Mittel

Der dritten Formel des Kategorischen Imperativs zufolge sollen die Menschen nicht nur als Selbstzweck gefördert werden, sondern gleichzeitig niemals als bloßes Mittel gebraucht werden. Deshalb wird im Folgenden interpretiert, was es heißt, als bloßes Mittel gebraucht zu werden. Dafür muss zuerst definiert werden, was es bedeutet, als Mittel gebraucht zu werden. Der Philosoph Samuel Kerstein, der sich unter anderem auf Kants Moralphilosophie fokussiert, definiert das folgendermaßen: „In sum, an agent uses another (or, equivalently, uses or treats another as a means) if and only if she intentionally does something to or with (some aspect of) the other in order to realize her end, and she intends the presence or participation of (some aspect of) the other to contribute to the end's

432 Kant 2008:67 [430]; Hervorhebungen im Zitat gesperrt.

realization."[433] Diese Definition ist in der Literatur weit verbreitet und anerkannt. Bevor die Frage gestellt wird, was es heißt, jemanden als *bloßes* Mittel zu benutzen, müssen einige Begrifflichkeiten geklärt werden.

Kleingeld macht einen Unterschied zwischen der *Behandlung* einer Person als Mittel und dem *Gebrauch* einer Person als Mittel. Kant verwendet meist den Begriff des Gebrauchens, der laut Kleingeld präziser ist als der Begriff des Behandelns.[434] Deshalb ist es ihr wichtig zu betonen, dass Kants dritte Formel des Kategorischen Imperativs es moralisch verbietet, eine andere Person als bloßes Mittel zu *gebrauchen*. Ein solcher Gebrauch kann nicht dadurch festgestellt werden, dass eine „(...) andere Person kausal zur Verwirklichung der Zwecke des Akteurs beiträgt."[435] Denn einen solchen Beitrag zu den Zwecken des Akteurs kann die andere Person auch von sich aus und freiwillig leisten. Zudem kann der Gebrauch einer anderen Person als Mittel nicht „(...) aus der Perspektive eines außenstehenden Erzählers (...)"[436] festgestellt werden, sondern die Handlung muss aus der Perspektive des Akteurs beschrieben werden. Nur wenn der Akteur seine Handlung danach ausrichtet, dass er eine andere Person zum Verwirklichen seiner Zwecke benutzen kann, dann gebraucht er sie als Mittel.[437]

In Bezug auf die Risikoauferlegung im Datenökosystem heißt das, Risikoauferlegende gebrauchen Risikoexponierte dann als Mittel, wenn sie ihnen das Risiko auferlegen, um ihre eigenen Zwecke zu verwirklichen, und dabei beabsichtigen, dass die Anwesenheit oder Teilnahme der Risikoexponierten zur Verwirklichung der Zwecke beiträgt. Im Kontext des digitalisierten Wohngeldantrages wären diese Zwecke zum Beispiel die Steigerung der Effizienz der Wohngeldbehörden sowie die erleichterte Antragsstellung. Die Antragstellenden werden also deshalb als Mittel gebraucht, weil es zum Verwirklichen der Zwecke notwendig ist, sie einem Risiko auszusetzen. Da die Antragstellenden als Risikoexponierte von diesen beiden Zwecken ebenfalls profitieren können, kann man davon ausgehen, dass dies

433 Kerstein 2013:58
434 Kleingeld 2024:117
435 Ebd.:118
436 Ebd.
437 Ebd.:119

auch ihre Zwecke sind, und somit geht die Benutzung als Mittel mit einem Gebrauch als Selbstzweck einher.

Kleingeld macht noch eine weitere Unterscheidung zwischen der Bewertung einer Handlung und der Bewertung der Maxime des Akteurs. Die Handlung selbst muss nur pflichtgemäß sein. Eine pflichtgemäße Handlung kann aus moralischen oder rechtlichen Gründen oder auch aus Eigennutz ausgeführt werden. Aber für die rechtliche Bewertung der Handlung ist nur relevant, ob die Handlung selber pflichtgemäß ist.[438] So kann zum Beispiel pflichtgemäß ein Versprechen eingehalten werden oder eine wohltätige Handlung ausgeführt werden – für die Bewertung der Handlung ist es erstmal gleichgültig, aus welchen Gründen es dazu kommt. Die Gründe für die Handlung werden relevant, wenn die Maxime des Akteurs bewertet werden soll. Eine Bewertung der Maxime des Akteurs kommt zum Beispiel dann ins Spiel, wenn festgestellt werden soll, ob andere Personen von dem Akteur als bloßes Mittel gebraucht wurden. Hier werden die zugrundeliegenden praktischen Überlegungen des Akteurs relevant. Hier macht es einen Unterschied, ob jemand eine ohnmächtige Joggerin reanimiert, um ihr Leben zu retten, weil man erkennt, dass man an die Hilfspflicht gebunden ist, oder ob man der Joggerin das Leben rettet, weil man sie noch braucht bzw. um danach als Held dazustehen. Im ersteren Fall handelt der Akteur nach der Maxime, anderen in Not zu helfen, „(...) die er sich aus Pflicht zu eigen gemacht hat."[439] Der Akteur behandelt die Joggerin nicht als Mittel und agiert moralisch richtig. Im zweiten Fall jedoch reanimiert der Akteur die Joggerin aus Eigennutz: „Indem er die unzulässige Maxime des Eigennutzes und nicht die moralisch gebotene der Wohltätigkeit annimmt, verstößt er gegen die Pflicht zur Wohltätigkeit (...)."[440] Das ist moralisch wertlos, weil der Akteur die Joggerin mit dieser Maxime bloß als Mittel gebraucht. Die Unzulässigkeit der Maxime macht allerdings nicht die Handlung unzulässig.[441]

Ebenso macht es einen Unterschied, ob der Wohngeldantrag digitalisiert wird, um den Prozess für die Antragsstellenden zu erleichtern, oder um möglichst viele personenbezogene Daten digital zu speichern, damit auch aus anderen Gründen, welche die Zwecke

438 Für die moralische Bewertung ist darüber hinaus auch die Motivation relevant.
439 Kleingeld 2024:118
440 Ebd.:126
441 Ebd.:126–128; Kant 2008:22 [398]

der Betroffenen nicht respektieren, darauf zugegriffen werden kann. Die erste Maxime wäre auch im Interesse der als Mittel gebrauchten Antragstellenden, aber die zweite Handlungsmaxime spricht von Eigeninteresse und ignoriert die Zwecke der Antragstellenden – damit wäre diese Maxime unzulässig. Was aber eine Handlung unzulässig macht, nämlich die Benutzung als bloßes Mittel, soll im Folgenden genau betrachtet werden.

5.2.3 Benutzung als bloßes Mittel

Kants dritte Formel des Kategorischen Imperativs verbietet nicht die Benutzung anderer Menschen als Mittel – diese liegt in der Natur jeder Kommunikation sowie Kooperation und lässt sich somit in einer Gesellschaft nicht vermeiden. Kant verbietet vielmehr das Benutzen anderer Menschen als *bloßes* Mittel und gebietet die Benutzung als Selbstzweck: „Handle so, daß du die Menschheit sowohl in deiner Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst."[442] Dieser Imperativ enthält also sowohl eine positive Verpflichtung, die Menschheit immer als Zweck zu verwenden, als auch eine negative Verpflichtung, die Menschheit niemals nur als Mittel zu verwenden. Dabei ist das Benutzen eines anderen Menschen als bloßes Mittel eine unrechtmäßige Instrumentalisierung.[443]

Was heißt es, jemanden bloß als Mittel zu benutzen? Kant nennt als Beispiel für eine Handlung, bei welcher man sein Gegenüber bloß als Mittel gebraucht, ein Versprechen, geliehenes Geld zurück zu zahlen, obwohl man weiß, dass man das Geld nicht zurückzahlen kann. Er begründet die moralische Verwerflichkeit einer solchen Handlung folgendermaßen: „Denn der, den ich durch ein solches Versprechen zu meinen Absichten brauchen will, kann unmöglich in meine Art, gegen ihn zu verfahren, einstimmen und also selbst den Zweck dieser Handlung enthalten."[444] Es gibt verschiedene Interpretationen dessen, was genau diese Unmöglichkeit, in die Behandlung einzustimmen, bedeutet: den Ansatz der Unmöglichkeit des Zweck-

442 Kant 2008: 65 [429]
443 Fahmy 2023:41
444 Kant 2008:66 [429–420]

Teilens, den Ansatz der unmöglichen Zustimmung, den Ansatz der tatsächlichen Zustimmung, und den pflichtbasierten Ansatz. Für meine Darstellung ist es wichtig alle vier Ansätze zu betrachten, weil sie sich ergänzen und ein umfassendes Verständnis dessen ermöglich, was es heißt jemanden bloß als Mittel zu gebrauchen.

Ansatz der Unmöglichkeit des Zweck-Teilens

Zum einen wird Kants Erklärung so verstanden, dass die Geld verleihende Person den Zweck der Geld leihenden Person, die verspricht das Geld zurück zu zahlen, obwohl sie das nicht kann, unmöglich teilen kann. So schreibt der Philosoph und Kant-Spezialist Allen W. Wood: „A false promise, because its end cannot be shared by the person to whom the promise is made, frustrates or circumvents that person's rational agency, and thereby shows disrespect for it."[445] In erster Linie wäre es praktisch irrational, den Zweck der Geld leihenden Person zu teilen, weil man als Geld verleihende Person dadurch finanziellen Verlust macht.[446] Es ist jedoch nicht ausgeschlossen, dass eine Geld verleihende Person einer möglichen Nichtrückzahlung zustimmt, wenn sie von den finanziellen Schwierigkeiten der leihenden Person weiß. Das Problem entsteht erst durch die Lüge. Denn wenn die Geld leihende Person ihre finanziellen Schwierigkeiten verschweigt und ein lügenhaftes Versprechen gibt, das Geld zurückzuzahlen, dann kann sich die Geld verleihende Person gar nicht zur Wahrheit verhalten und der Nichtrückzahlung zustimmen. Durch die Lüge hat die Geld verleihende Person die Wahl, ob sie zu dem Zweck beitragen möchte.[447] Außerdem kann einem lügenhaften Versprechen nicht zugestimmt werden, weil daraus Beliebigkeit folgt – aus dem Gesagten folgt dann nichts mehr, da unklar ist, ob das Gesagte oder etwas anderes gemeint ist. Somit ist es logisch unmöglich den Zweck der lügenden Person zu teilen.[448]

In der Literatur werden vier verschiedene sich ergänzende Gründe diskutiert für die Unmöglichkeit, den Zweck der einen behandelnden Person zu teilen. So kann es erstens logisch unmöglich sein,

445 Wood 1999:153
446 Kerstein 2009:170
447 Korsgaard 1996:139
448 Hill 2002:69f.

den Zweck zu teilen[449]; zweitens kann die Wahl, ob man zu dem Zweck beitragen möchte, verwehrt sein[450], obwohl man einen legitimen Anspruch darauf hat, an dieser Entscheidung nicht gehindert zu werden, da durch den Zweck den eigenen Interessen geschadet wird[451]; drittens kann es praktisch irrational sein, den Zweck zu teilen (hypothetischer Imperativ)[452]; viertens kann der Zweck nicht geteilt werden, wenn er unvereinbar ist mit der Förderung des Zwecks der Achtung der Menschlichkeit, da dies ein Zweck ist, zu dem jeder Mensch rational verpflichtet ist[453].

Die Moralphilosophin Christine Korsgaard, die einflussreiche Beiträge zur Neuinterpretation der Philosophie von Kant verfasst hat, meint es ist dann unmöglich den Zweck der einen behandelnden Person zu teilen, wenn sie einen daran hindert, zu wählen, ob man zur Verwirklichung dieses Zweckes beitragen will oder nicht.[454] Die feministische Moralphilosophin Lina Papadaki, die ebenfalls auf Kant spezialisiert ist, ergänzt Korsgaards Grund der fehlenden Wahl um den notwendigen legitimen Anspruch darauf, da es Fälle geben kann, in welchen man keinen legitimen Anspruch hat zu wählen, ob man als bloßes Mittel genutzt werden möchte oder nicht. So zum Beispiel, wenn eine Kundin ein Getränk bei einem Kellner bestellt, um heimlich einen besseren Blick auf den attraktiven Barkeeper erhaschen zu können. Solange man das übergeordnete Ziel der handelnden Person teilen kann, habe man keinen legitimen Anspruch darauf, zu entscheiden, ob man zur Verwirklichung des Ziels der Person beitragen will oder nicht. Da der Kellner dem übergeordneten Ziel der Kundin, bei ihm ein Getränk zu bestellen, bei der Annahme seines Jobs zugestimmt hat, hat er keinen legitimen Anspruch zu entscheiden, ob er zur Verwirklichung des zusätzlichen Ziels der Kundin, einen besseren Blick auf den Barkeeper zu bekommen, beitragen will oder nicht. Denn die Kundin schadet den Interessen des Kellners nicht, indem sie ein Getränk bei ihm bestellt. Wenn hingegen eine Geld leihende Person der Geld verleihenden

449 Hill 2002:69f.
450 Korsgaard 1996:139
451 Papadaki 2016:80
452 Kerstein 2009:170
453 Papadaki 2016:92f.
454 Korsgaard 1996:139

Person verspricht das Geld zurück zu zahlen, obwohl sie das nicht kann, würde der Handlungszweck den Interessen der Geld verleihenden Person schaden. Somit ist das Wissen der Geld verleihenden Person über das Ziel der Geld leihenden Person moralisch relevant. In dem Fall hat die Geld verleihende Person einen legitimen Anspruch darauf, über das Ziel der Geld leihenden Person informiert zu werden.[455]

Papadaki fügt außerdem die Unvereinbarkeit mit der Förderung des Zwecks der Achtung der Menschlichkeit als Grund hinzu, aus welchem die Zwecke der Benutzenden nicht geteilt werden können. Sie argumentiert, dass über die drei anderen Gründe zum Beispiel bei einer Person, die sich freiwillig als Sklavin behandeln lässt, nicht festgestellt werden könnte, dass diese Person als bloßes Mittel benutzt wird. Mit Papadakis Ergänzung des rationalen Zwangs zur Förderung der Achtung der Menschlichkeit ist dies jedoch möglich. Denn auch eine Person, die keine logischen oder rationalen Fehler macht, und wählt, sich als Sklavin behandeln zu lassen, darf laut Papadaki nicht als Sklavin behandelt werden, da dies gegen die Pflicht zur Achtung der Menschlichkeit verstößt.

Es ist also (erstens) logisch unmöglich, Geld zu verleihen und dabei den Zweck zu teilen, dass das Geld nicht zurückgegeben wird – das wäre dann eine Geldschenkung oder die Bereitschaft, das Risiko einzugehen, dass die leihende Person das Geld eventuell nicht wird zurückzahlen können. Es ist hingegen nicht logisch unmöglich, den Zweck von jemandem zu teilen, der einem ein Risiko auferlegt, da der Zweck der Risikoauferlegung einer sein kann, von dem man selbst profitiert, wie beim Wohngeldantrag. Denn dann gehört man zur Partei der durch das Risiko Begünstigten und hat einen Grund das Risiko für die Begünstigung in Kauf zu nehmen.[456] In gleicher Weise ist es (drittens) auch nicht praktisch irrational, den Zweck derjenigen zu teilen, die einem ein Risiko auferlegen, solange man selbst auch davon profitiert.[457] Da der Zweck einer Risikoauferlegung (viertens) mit der Förderung des Zwecks der Achtung der Menschlichkeit vereinbar sein kann, solange er dem Vorteil aller Beteiligten dient, ist auch dies kein Grund, den Zweck der

455 Papadaki 2016:79f.
456 Hermansson, Hansson 2007:130
457 Solange dieser Profit selbst wiederum praktisch rational ist.

Risikoauferlegenden nicht zu teilen. Jedoch ist bei einer solchen Risikoauferlegung (zweitens) die Wahl verwehrt, ob man zu dem Zweck beitragen möchte, während man aber einen legitimen Anspruch darauf hat, an dieser Entscheidung nicht gehindert zu werden, da man durch die Risikoauferlegung zu Schaden kommen kann. Der Ansatz der Unmöglichkeit des Zweck-Teilens würde also nach der Auslegung von Korsgaard samt Ergänzung von Papadaki dafürsprechen, dass Risikoexponierte durch das Auferlegen eines Risikos ohne Wahlmöglichkeit als bloßes Mittel gebraucht werden, selbst wenn sie von den Zwecken der Risikoauferlegung profitieren.

Ansatz der unmöglichen Zustimmung

Der Ethikerin und politischen Philosophin Onora O'Neill zufolge kann die Frage danach, ob man den Zweck der Behandlung durch andere teilen kann, nur dabei helfen festzustellen, ob man als Person behandelt wird. Um festzustellen, ob man (als bloßes Mittel) gebraucht wird, müsse man die Frage stellen, ob man dem Gebrauch zustimmen könne. Wenn man aber zum Beispiel gezwungen oder getäuscht wird, ist der Widerspruch und damit die echte Zustimmung ausgeschlossen. Deshalb würden bei Zwang oder Täuschung andere Menschen als bloße Instrumente in den Projekten der Handelnden benutzt.[458] O'Neill schreibt: „Victims cannot agree with a coercer's fundamental principle or maxim, which denies them the choice between consent and dissent, and further cannot share a coercer's ends. (…). Those who are either deceived or coerced are then *both* used *and* not treated as persons. "[459] Das heißt, wenn Menschen getäuscht werden, damit sie ein Risiko akzeptieren, oder wenn sie dazu gezwungen werden, ist das, laut O'Neills Ansatz der unmöglichen Zustimmung, als Gebrauch als bloßes Mittel zu klassifizieren. Wenn unter Zwang auch verstanden wird, dass Risiken unfreiwillig akzeptiert werden, weil es keine gute Alternative gibt[460], dann habe ich oben gezeigt, dass dies im Kontext von Risikoauferlegung oft vorkommt.

458 O'Neill 1989:111
459 Ebd.:113
460 Wolff, De-Shalit 2007:66

Es gibt aber eine Möglichkeit, den Gebrauch als bloßes Mittel zu verhindern: Die Voraussetzung dafür, dass die benutzte Person den Zweck der Handelnden teilen und somit als Person behandelt werden kann, ist laut O'Neill die Möglichkeit der Zustimmung zu der Benutzung: „To treat others as persons we must allow them the *possibility* either of consenting to or of dissenting from what is proposed."[461] Es sei nicht moralisch bedeutsam, welcher Behandlung ein Mensch tatsächlich oder hypothetisch zustimmen würde, sondern es gehe darum, die Zustimmung oder Ablehnung des Gebrauchs zu ermöglichen.

Papadaki kritisiert, wie auch bezüglich der Unmöglichkeit des Zweck-Teilens ausgeführt, dass eine Person zustimmen könnte, als Sklavin benutzt zu werden und dann laut O'Neills Ansatz nicht als bloßes Mittel benutzt würde. Papadaki besteht darauf, dass jeder Mensch rational verpflichtet ist, den Zweck der Achtung der Menschlichkeit zu fördern.[462] Die Zustimmung, als Sklavin benutzt zu werden, würde bedeuten, dass man auf die Förderung des Zwecks der Achtung der Menschlichkeit verzichtet, zu dem man aber rational verpflichtet ist.[463] Sie schlägt auch hier vor, den Ansatz der unmöglichen Zustimmung von O'Neill zu ergänzen um die Pflicht zur Förderung der Achtung der Menschlichkeit. Die Differenz zwischen O'Neill und Papadaki scheint die Frage zu betreffen, ob die Möglichkeit der Zustimmung hinreichend ist (O'Neill) oder ob dasjenige, wozu man zustimmt, eine bestimmte, objektive normative Qualität aufweisen muss (Papadaki).

Ansatz der tatsächlichen Zustimmung

Kleingeld macht ein Angebot für eine ähnliche Auslegung, wann eine Person bloß als Mittel gebraucht wird:

> „Ein Akteur gebraucht eine andere Person dann und nur dann bloß als Mittel, wenn (1) er diese andere Person *als Mittel* im Dienste der Verwirklichung seiner Zwecke *gebraucht*, (2) ohne diesen Gebrauch moralisch-prinzipiell von der *Zustimmung des* anderen *abhängig zu* machen;

461 O'Neill 1989:110
462 Papadaki 2016:92f.
463 Ebd.:93

> wobei (3) mit „Zustimmung" die echte *tatsächliche* Zustimmung des anderen, in einer bestimmten Weise als Mittel zum Zweck des Akteurs gebraucht zu werden, gemeint ist."[464]

Die erste Bedingung stellt sicher, dass eine Person als Mittel zu einem Zweck gebraucht wird, der nicht ihr eigener ist, bevor sie als bloßes Mittel gebraucht werden kann.[465] Kleingeld erklärt die zweite Bedingung folgendermaßen: „Angesichts des moralischen Status von Personen als Selbstzwecke(...) ist es für eine Person moralisch unzulässig, eine andere als Mittel zu gebrauchen ohne ihre Zustimmung."[466] Man kann andere Personen trotzdem als Mittel benutzen, wenn man diese Benutzung von ihrer Zustimmung abhängig macht. Aber man gebraucht eine andere Person nur dann *bloß* als Mittel, wenn man diesen Gebrauch nicht von deren Zustimmung abhängig macht. Ihre dritte Bedingung begründet Kleingeld mit Hilfe der Analyse von Kants Beispiel des falschen Versprechens, da es sich ihrer Meinung nach hierbei um einen Fall handelt, bei dem „(...) die *tatsächliche* Zustimmung fehlt."[467] Der Gegenstand dieser tatsächlichen Zustimmung ist nicht die Maxime der Gebrauchenden, sondern der Gebrauch selbst.

Bei einer Risikoauferlegung wird die risikoexponierte Person als Mittel zu einem ihr äußerlichen Zweck gebraucht, sobald der Wert des Zwecks der Risikoauferlegung den Wert des Vorteils für die Risikoexponierten überragt. Wenn zum Beispiel den Wohngeldantragstellenden ein Risiko auferlegt wird, um unter anderem auch den Wohngeldbehörden und dem Staat zu nutzen, werden die Antragstellenden als Mittel gebraucht zur Verwirklichung der Zwecke anderer. Laut Kleingeld werden die Antragstellenden jedoch erst dann als bloßes Mittel gebraucht, wenn die Risikoauferlegung nicht von ihrer Zustimmung abhängig gemacht wird und wenn ihre tatsächliche Zustimmung zu dem Gebrauch fehlt. Hier fragt sich, was tatsächliche Zustimmung bedeutet. Denn im Prinzip stimmen die Antragstellenden beim Stellen des Wohngeldantrags zu, dass ihre Daten für die Wohngeldstatistik weitergegeben werden dürfen, aber sie haben keine Alternative, wenn sie den Antrag stellen wollen. Es

464 Kleingeld 2024:116
465 Ebd.:117
466 Ebd.:120
467 Ebd.:121

ist allerdings fraglich, ob die Verfügbarkeit mehrerer Alternativen normativ ausschlaggebend ist. Vielmehr ist entscheidend, ob die richtige Option verfügbar ist – nicht ob es viele Optionen gibt, die alle inakzeptabel sein könnten.

Kleingeld meint, dass eine Zustimmung nur unter bestimmten Bedingungen als *tatsächliche* Zustimmung gewertet werden darf – sie nennt den Ausschluss von Täuschung, das Vorliegen geistiger Reife und die Verfügung über relevante und verständliche Informationen als Bedingungen: „Die Person muss wissen, welchem Zweck sie dienen soll, wie sie eingesetzt werden soll, was von ihr verlangt wird und so weiter."[468] An dieser Stelle verweist sie auf eine ausführlichere Diskussion, die geführt werden müsste, um weitere Details zur Zustimmung zu klären. Vorerst ist jedoch klar, dass Kleingeld O'Neill zustimmen würde, dass bei Täuschung keine tatsächliche Zustimmung möglich ist. Sie fügt jedoch hinzu, dass relevante und verständliche Informationen über den Gebrauch notwendig sind, um diesem tatsächlich zustimmen zu können.

Des Weiteren stellt Kleingeld, ganz ähnlich wie Papadaki, fest, „(...) dass diejenigen, die ihre Zustimmung *geben*, ihrerseits auch bestimmte(...) moralische(...) Anforderungen erfüllen müssen."[469] Da man sich laut Kant auch selbst nicht bloß als Mittel gebrauchen darf, ist es moralisch nicht möglich, jedem Gebrauch der eigenen Person für die Zwecke von anderen zuzustimmen. Diese Bedingung sollte laut Kleingeld auch von den Akteuren beachtet werden, die um Zustimmung bitten: „(...) denn er darf andere nicht um deren Zustimmung dazu bitten, auf erniedrigende oder anderweitig moralisch unzulässige Weise benutzt zu werden."[470] Das funktioniert allerdings nicht anders herum: „Nach Kants Ansicht besteht eine Pflicht zur Wohltätigkeit, sodass man unter bestimmten Umständen durchaus moralisch verpflichtet sein kann, seine Zustimmung zu geben."[471] Diese moralische Verpflichtung zuzustimmen reicht aber nicht aus, um andere Personen ungefragt als Mittel zu einem wohltätigen Zweck einzusetzen. Wenn die Personen nicht nach ihrer Zustimmung gefragt werden, werden sie auch dann bloß als Mittel benutzt.

468 Kleingeld 2024:123
469 Ebd.
470 Ebd.
471 Ebd.

Pflichtbasierter Ansatz

Die Philosophin Melissa Seymour Fahmy, die sich auf die Erforschung der Kantischen Ethik spezialisiert hat, ist der Meinung, dass die Ansätze des Zweck-Teilens und der unmöglichen bzw. tatsächlichen Zustimmung nicht ausreichen, und sie stellt als Alternative einen pflichtbasierten Ansatz auf: „X uses Y merely as a means if and only if (1) X uses Y as a means and (2) X's use of Y violates either a consent-sensitive duty that X owes to Y or a consent-insensitive duty that X owes to Y."[472] Sie unterscheidet zwischen zwei Arten von Pflichten, die wir haben können, wenn wir andere Leute benutzen: Zustimmungsbedürftige Pflichten (consent-sensitive duties) und zustimmungsunempfindliche Pflichten (consent-insensitive duties): „Whereas consent-sensitive duties have a consent condition – it is wrong to do x unless you have consent from Y – consent-insensitive duties have no consent condition. Negative consent-insensitive duties prohibit forms of conduct regardless of whether the conduct has been consented to."[473] So ist es zum Beispiel falsch, das Eigentum eines anderen zu benutzen, außer der andere stimmt meiner Benutzung zu – diese zustimmungsbedürftige Pflicht, das Eigentum anderer nicht zu stehlen, kann also durch die besitzende Person aufgehoben werden. Allerdings scheint es auch bei meiner Zustimmung nicht zulässig zu sein, mir tödlichen Schaden zuzufügen, da ich selbst die Pflicht habe, mich nicht zu verletzen – deshalb ist es eine zustimmungsunempfindliche Pflicht, andere nicht zu töten. Gleichermaßen ist es laut Fahmy unzulässig, sich selbst nicht zu respektieren.[474] Deshalb können wir andere nicht von ihren Respektpflichten uns gegenüber entbinden, selbst wenn ich einer respektlosen Behandlung zustimme, besteht für die handelnde Person weiterhin die Pflicht, mich als würdevollen Menschen zu respektieren. Das heißt, die Pflicht zum Respekt ist zustimmungsunempfindlich. Wenn wir solche zustimmungsunempfindlichen Pflichten verletzen – mit oder ohne die Zustimmung der Benutzten – tun wir ihnen Unrecht. Denn wenn wir die behandelten Menschen dabei als Mittel benutzten, benutzen wir sie als bloßes Mittel.

472 Fahmy 2023:55
473 Ebd.:49
474 Ebd.:50

Diejenigen, die über die Risikoauferlegung entscheiden, benutzen Risikoexponierte dann *bloß* als Mittel, wenn sie Risikoexponierte als Mittel verwenden, und wenn die Verwendung der Risikoexponierten entweder eine zustimmungsunempfindliche Pflicht oder eine zustimmungsbedürftige Pflicht verletzt. Es ist jedoch deutlich, dass es keine zustimmungsunempfindliche Pflicht ist, anderen kein Risiko aufzuerlegen. Denn eine Person kann sich selbst Risiken aussetzen – dadurch wird keine Pflicht verletzt, solange sich die Person dabei selbst respektiert. Es würde sich dann um eine zustimmungsbedürftige Pflicht handeln, wenn es falsch wäre, andere einem Risiko auszusetzen, ohne ihre Zustimmung zu haben. Genau das habe ich in den vorherigen Abschnitten mit Hilfe von Korsgaard, O'Neill und Kleingeld begründet:

- Wenn die Wahl, ob man zu dem Zweck beitragen möchte, verwehrt ist[475], obwohl man einen legitimen Anspruch darauf hat, an dieser Entscheidung nicht gehindert zu werden, da durch diese zweckhafte Handlung den eigenen Interessen geschadet wird[476], oder
- es unmöglich ist zuzustimmen, weil man zum Beispiel zur Akzeptanz eines Risikos gezwungen wird[477], oder
- wenn man nicht nach seiner tatsächlichen Zustimmung gefragt wird, ohne Täuschung und mit relevanten Informationen[478],

dann wird man als bloßes Mittel gebraucht, und das ist laut Kants dritter Formel des Kategorischen Imperativs verboten.[479] Deshalb gehe ich davon aus, dass es falsch ist, andere ohne deren Zustimmung einem Risiko auszusetzen und dass es dementsprechend eine zustimmungsbedürftige Pflicht ist, andere keinem Risiko auszusetzen. Das heißt, solange das Risiko keine Zustimmung zu tödlichem Schaden oder zu Respektlosigkeit von den Risikoexponierten erfordert, kann die Zustimmung der Risikoexponierten die Risikoauferlegenden von der Pflicht befreien, den Risikoexponierten kein Risiko aufzuerlegen.

475 Korsgaard 1996:139
476 Papadaki 2016:80
477 O'Neill 1989:113
478 Kleingeld 2024:116
479 Kant 2008: 65 [429]

Insgesamt kann man sagen, dass diejenigen, die über die Risikoauferlegung entscheiden, Risikoexponierte ohne deren Zustimmung *bloß* als Mittel gebrauchen, wenn sie Risikoexponierte als Mittel gebrauchen und der Gebrauch der Risikoexponierten eine Pflicht verletzt. Diese von Kant verbotene Benutzung der Risikoexponierten als bloßes Mittel kann jedoch laut Fahmy abgewendet werden, wenn ihre gültige Zustimmung eingeholt wird: „(…) when I use the phrase valid consent, it should be understood to mean consent given voluntarily, by a decisionally competent agent, with understanding appropriate for the context."[480] Fahmys drei Bedingungen für gültige Zustimmung sind Kleingelds Bedingungen für tatsächliche Zustimmung sehr ähnlich. Die Zustimmung soll freiwillig sein und dementsprechend ist Täuschung ausgeschlossen, die zustimmenden Akteure sollen entscheidungsfähig sein bzw. geistige Reife vorweisen können und sie sollen über ein dem Kontext angemessenes Verständnis verfügen, also über relevante und verständliche Informationen verfügen.[481] Auch Korsgaard meint eine Zustimmung ist nur dann möglich, wenn die betreffende Person Wissen hat über die Vorgänge und ein gewisses Maß an Einfluss auf den Ablauf.[482] Die Zustimmung ist laut Fahmy, genau wie bei O'Neill, ungültig, wenn sie beispielsweise unter Zwang oder Täuschung gegeben wird.[483] Zudem kann die Zustimmung zu einer zustimmungsunempfindlichen Pflicht diese Pflicht nicht aufheben. Denn es gelten genau wie bei Papadaki und Kleingeld moralische Anforderungen sowohl an die Akteure, die um Zustimmung bitten als auch diejenigen, die ihre Zustimmung geben. Denn ein Gebrauch auf respektlose, erniedrigende oder anderweitig moralisch unzulässige Weise, welche die Förderung des Zwecks der Achtung der Menschlichkeit vernachlässigt, ist nicht zulässig.[484]

480 Fahmy 2023:48
481 Kleingeld 2024:123; Fahmy 2023:48
482 Korsgaard 1996:139
483 Fahmy 2023:60, Fußnote 18
484 Papadaki 2016:93; Fahmy 2023:50; Kleingeld 2024:123

5.2.4 Zwischenfazit

Ich denke, dass sich diese vier Ansätze nicht widersprechen, wie die Autorinnen zum Teil behaupten, sondern dass sie sich ergänzen – Fahmys pflichtbasierter Ansatz bildet dabei einen Rahmen für die drei anderen Ansätze. Denn wir haben die zustimmungsunempfindliche Pflicht, unser Gegenüber mit Respekt zu behandeln[485] und durch unser Handeln die Achtung der Menschlichkeit zu fördern[486], sowie andere nicht auf erniedrigende oder anderweitig moralisch unzulässige Weise zu benutzen[487]. Eine zustimmungsbedürftige Pflicht ist hingegen, andere nicht als Mittel zu gebrauchen, denn von dieser Pflicht können die Gebrauchenden durch die Zustimmung der Gebrauchten befreit werden.[488] Bei der Verletzung der zustimmungsbedürftigen Pflicht, welche die Gebrauchenden den Gebrauchten schuldet, werden die Gebrauchten als bloßes Mittel benutzt.[489] Diese Pflicht wird dann verletzt, wenn die Betroffenen ohne ihre moralisch qualifizierte Zustimmung als Mittel gebraucht werden.[490] Dabei geht es den Autorinnen um die Möglichkeit der Zustimmung bzw. der Ablehnung des Gebrauchs[491], um tatsächliche Zustimmung[492] beziehungsweise um valide Zustimmung[493]. Alle drei Ansätze meinen damit jedoch gleichermaßen, dass die Zustimmung freiwillig sein soll und nicht durch Zwang oder Täuschung herbeigeführt werden darf, die zustimmenden Akteure sollen geistig fähig sein, eine echte Zustimmung geben zu können, und sie sollen über ein dem Kontext angemessenes Verständnis verfügen, sprich auf relevante und verständliche Informationen Zugriff haben.[494]

Auch die Pflicht, unserem Gegenüber die Wahl zu lassen, ob es von uns als bloßes Mittel gebraucht werden möchte[495], kann als

485 Fahmy 2023:50
486 Papadaki 2016:92f.
487 Kleingeld 2024:123
488 O'Neill 1989:111; Kleingeld 2024:116; Fahmy 2023:55
489 Fahmy 2023:55
490 Kleingeld 2024:116; O'Neill 1989:111
491 O'Neill 1989:110
492 Kleingeld 2024:116
493 Fahmy 2023:48
494 Kleingeld 2024:123; Fahmy 2023:48, 60, Fußnote 18; O'Neill 1989:111
495 Korsgaard 1996:139

zustimmungsbedürftig gelten, wenn man davon ausgeht, dass die Zustimmung des Gegenübers uns von der Pflicht befreien kann. In diesem Fall müsste man zustimmen, nicht mehr die Wahl zu haben, ob man zu dem Zweck beitragen möchte, obwohl man einen legitimen Anspruch darauf hat, an dieser Entscheidung nicht gehindert zu werden.[496] Hier hängt der Verzicht auf eine Wahl davon ab, zu welchem Zweck man beiträgt. Wie oben erörtert, muss die Risikoauferlegung den Zweck der Achtung der Menschlichkeit fördern, damit man dieser zustimmen kann.[497] Außerdem kann der Verzicht auf die Wahl attraktiver sein, wenn man selber von dem Risiko profitiert und somit zur Partei der durch das Risiko Begünstigten gehört[498] – dadurch hätte man einen Grund das Risiko für die Begünstigung in Kauf zu nehmen. Allerdings ist den Begünstigten dann trotz ihres Profits immer noch die Wahl verwehrt, ob sie zu dem Zweck der Risikoauferlegung beitragen möchten[499], obwohl man einen legitimen Anspruch darauf hat, an dieser Entscheidung nicht gehindert zu werden[500], da den eigenen Interessen durch die Risikoauferlegung neben der Begünstigung auch geschadet wird. Wie unter dem *Ansatz der Unmöglichkeit des Zweck-Teilens* festgestellt, werden Risikoexponierte durch das Auferlegen eines Risikos ohne Wahlmöglichkeit als bloßes Mittel genutzt, selbst wenn sie von den Zwecken der Risikoauferlegung profitieren.

Kants dritte Formel des Kategorischen Imperativs kann demnach so interpretiert werden, dass es moralisch verboten ist, andere als Mittel zu gebrauchen, ohne dass ihre Zustimmung in moralisch relevanter Weise eingeholt wird – sei es,

- weil diese Zustimmung praktisch irrational[501] oder logisch unmöglich[502] ist,
- weil es durch Zwang oder Täuschung keine Möglichkeit zur Zustimmung gibt[503] oder die Wahl anderweitig verwehrt ist[504], obwohl man einen legitimen Anspruch darauf hat[505],

496 Papadaki 2016:80
497 Ebd.:92f.
498 Hermansson, Hansson 2007:130
499 Korsgaard 1996:139
500 Papadaki 2016:80
501 Kerstein 2009:170
502 Hill 2002:69f.

- weil dadurch die zustimmungsunempfindliche Pflicht, unser Gegenüber mit Respekt zu behandeln verletzt wird[506], der Zweck der Achtung der Menschlichkeit nicht gefördert wird[507] oder die Zustimmung zu erniedrigendem oder anderweitig moralisch unzulässigem Gebrauch verlangt wird[508],
- weil die tatsächliche Zustimmung ohne geistige Reife sowie relevante und verständliche Informationen nicht möglich ist[509] oder weil die valide Zustimmung ohne Freiwilligkeit, Entscheidungsfähigkeit und ein dem Kontext angemessenes Verständnis nicht möglich ist[510].

Diese Zusammenfassung macht deutlich, dass die moralische Zulässigkeit von Risikoauferlegung maßgeblich davon abhängt, ob die Risikoexponierten informiert und freiwillig zustimmen können. So lange die Zustimmung nicht praktisch irrational oder logisch unmöglich ist und die Menschlichkeit achtet, gebietet Kants dritte Formel des Kategorischen Imperativs, den Risikoexponierten die ungezwungene Wahl zu lassen, ob sie das Risiko tragen wollen und sie mit relevanten und verständlichen Informationen auszustatten, damit sie entscheidungsfähig sind. In Bezug auf die Auferlegung von Risiken im Datenökosystem bedeutet dies, dass es der Kategorische Imperativ verbietet, anderen ein Risiko aufzuerlegen, ohne ihnen die Wahl zu lassen und ohne sie darüber zu informieren, was sie wählen. Wenn also Kants dritte Formel des Kategorischen Imperativs eine Regel für ein gerechtes soziales System der Risikobereitschaft[511] ist, dann sind nur die Arten von risikobehafteten Aktivitäten zugelassen, welche die Risikoexponierten über das Risiko informieren und ihnen die Wahl lassen, ob sie es tragen wollen. Diese Bereitschaft das Risiko zu akzeptieren, hängt dabei davon ab, ob die Risikoexponierten dabei auch darin befördert werden, ihre eigenen qualifizierten

503 O'Neill 1989:111
504 Korsgaard 1996:139
505 Papadaki 2016:80
506 Fahmy 2023:50
507 Papadaki 2016:92f.
508 Ebd.
509 Ebd.:123
510 Fahmy 2023:48
511 Hansson 2003:305

Zwecke zu verfolgen, ob sie selbst von dem Risiko profitieren und ob der Zweck des Risikos die Achtung der Menschlichkeit fördert.

5.2.5 Umsetzungsschwierigkeiten und Lösungsansätze

In der Praxis ergeben sich Probleme bei der Umsetzung von Bedingungen für ein gerechtes soziales System der Risikobereitschaft, die sich aus Kants dritter Formel des Kategorischen Imperativs ableiten lassen. So ist zum Beispiel die Möglichkeit zum Widerspruch in einem Datenökosystem in der öffentlichen Verwaltung sehr schwer umzusetzen, wenn das gesamte Verwaltungssystem digitalisiert wird. Dann müsste es die Wahl geben zwischen einer analogen und einer digitalen Antragsstellung und -bearbeitung. Wenn die Verwaltung eine solche Wahlmöglichkeit anbieten müsste, um den Bürgern Kontrolle zu ermöglichen, dann wäre sie wieder nicht in der Lage, die Anträge der Bürger zeitnah zu bearbeiten und die Digitalisierung hätte die Umstände nicht verbessert. Laut Luhmann „(...) liegt [es] auf der Hand, daß nicht alle an allen Entscheidungen beteiligt werden können (...).“[512] Vielmehr sei die Betroffenheit derjenigen, die nicht mitentscheiden dürfen, eine Mobilisierungsressource und sie habe deshalb eine positive Funktion für die Selbstbeobachtung der Gesellschaft.[513]

Da die freie Wahl des Risikos eines jeden Bürgers also nicht praktikabel ist, könnte der Staat paternalistisch vorgehen und im Interesse der risikoexponierten Bürger handeln. Fahmy ist der Meinung, dass nur Kinder und Betrunkene paternalistisch behandelt werden sollten: „If a drunk person is like a child, then it makes sense to treat the drunk person in a manner comparable to how we should treat children: paternalistically.”[514] Ich gehe allerdings davon aus, dass die Bürger:innen bei der Nutzung von Verwaltungsservices keine Kinder mehr und nicht betrunken sind. Jedoch stellt sich die Frage, ob deutsche Bürger einen legitimen Anspruch auf diese Wahl haben.[515] Es wäre wohl nicht legitim, wenn jeder einzelne Bürger einen Anspruch darauf erheben würde mitzubestimmen, wie das deutsche

512 Luhmann 1990:152–153
513 Ebd.
514 Fahmy 2023:56f.
515 Papadaki 2016:80

Verwaltungssystem funktioniert – es ist ja schon schwer genug, sich innerhalb einer Regierung darauf zu einigen. Außerdem gibt es ein Gesetz (OZG), das die Digitalisierung der Verwaltung vorschreibt – die Umsetzung steht also unmittelbar bevor. Allerdings gibt es Gestaltungsspielraum, was die Einflussmöglichkeiten der Bürger innerhalb des digitalisierten Systems angeht. Die Gestaltungsaufgabe fällt dem Staat zu, dieser muss aber im Interesse der Bürger handeln.

Woher aber soll der Staat wissen, was im Interesse aller Bürger liegt? Tatsächlich ist das in diesem Fall einheitlich. Kein Bürger möchte zu Schaden kommen, indem er die Kontrolle über seine Daten verliert und diese womöglich durch Unbefugte zweckentfremdet werden oder indem er verzerrten Ergebnisse ausgesetzt ist. Dabei befinden sich die staatlichen Entscheidenden quasi hinter einem „Schleier des Nichtwissens“[516], da auch sie als Staatsbürger potenziell von einem solchen Schaden betroffen sein könnten. Man könnte argumentieren, dass manche Personen einem höheren Schadensrisiko im Sinne von Diskriminierung ausgesetzt sind, zum Beispiel auf Grund ihrer Hautfarbe, Religion oder einer psychischen Erkrankung. Aber da alleine der Kontrollverlust und die Unwissenheit, die jeden Bürger gleichermaßen treffen können, schon einen Schaden darstellen, kann zumindest davon jeder Bürger betroffen sein.

Dadurch, dass das Recht der Antragstellenden, keinem Risiko ausgesetzt zu werden, durch eine Risikoauferlegung außer Kraft gesetzt wird, steht den Risikoexponierten laut Hansson Kompensation zu.[517] Er argumentiert, dass hierdurch zusätzliche Pflichten entstehen für diejenigen, die das Risiko auferlegen: „(...) such as obligations to compensate, to inform, and to reduce the risk whenever possible.“[518] Mit Hilfe des Drei-Parteien-Modells ließ sich jedoch erkennen, dass Informieren nicht ausreicht, wenn die Betroffenen daraufhin keine freie Wahl haben, ob sie das Risiko akzeptieren oder nicht.[519] Nur wenn die Betroffenen nicht gezwungen sind, kann diese Risikoaussetzung Teil eines gerechten sozialen Systems der Risikobereitschaft

516 Rawls 1975:159

517 Wie oben nehme ich an, dass ein solches Recht nur aus einer moralischen Pflicht, niemandem einem Risiko auszusetzen, abgeleitet werden kann.

518 Hansson 2003:303

519 Hermansson, Hansson 2007:142

sein, das den Risikoexponierten zum Vorteil gereicht.[520] Wenn allerdings eine Maßnahme, wie die Digitalisierung von Verwaltungsleistungen, nur flächendeckend möglich ist, sie aber gesellschaftlich geboten ist, ist sie zumutbar.

Fahmy ist der Meinung, dass es ohne Möglichkeit der Zustimmung darum geht, die Risikoexponierten vor Schaden zu schützen und möglichst ihre Handlungsfähigkeit wiederherzustellen: „When it comes to persons with impaired agency (agents who cannot give valid consent), treating them as ends in themselves seems to require that we provide care that aims to protect them from immediate harm and restore their agency (when possible).“[521] Bei dem Gebrauch von Personen, die selbst nicht über diesen Gebrauch entscheiden können, sollte es das Ziel sein, die Handlungsfähigkeit der betroffenen Person wiederherzustellen, möglichst ohne dass sie zu Schaden kommt. Dabei spricht Fahmy allerdings von gebrauchten Personen, die nicht bei Bewusstsein sind: „The agent who performs CPR on the unconscious jogger acts in accordance with the consent-insensitive duty to come to the aid of impaired agents. And though she uses the unconscious jogger as a means to her end of rendering aid, she violates neither a consent-sensitive nor a consent-insensitive duty. “[522] Im Gegensatz zu der Joggerin, die wiederbelebt wird, sind die Bürger:innen, welche die Verwaltungsservices in Anspruch nehmen, alle bei Bewusstsein. Die Joggerin, die nicht bei Bewusstsein ist, kann keine Kontrolle übernehmen, bevor ihr geholfen wird – das Helfen ist deshalb zulässig, weil es ihre Fähigkeit, die Kontrolle zu übernehmen, wieder herstellt. Die Präferenzen der bewusstlosen Joggerin sind uns nicht bekannt, aber unsere Pflichten, uns selbst und anderen gegenüber, sind uns bekannt.

In ähnlicher Weise sollte auch die Handlungsfähigkeit der Datenquellen und Datennutzenden wiederhergestellt werden: Es ist zwar logistisch nicht möglich, jeder einzelnen betroffenen Person die Wahl zu lassen, ob die Verwaltung digitalisiert wird und ob dementsprechend ihre Daten gesammelt, gespeichert und (mit welchem Algorithmus) verarbeitet werden dürfen, aber es ist möglich, innerhalb dieses Datenökosystems den gebrauchten Akteuren so viel Kontrolle

520 Hansson 2003:305
521 Fahmy 2023:57
522 Ebd.

wie möglich zu übertragen. Zum Beispiel könnte es im Rahmen einer digitalen Cloud individuelle digitale „Wallets" geben, die von den jeweiligen Datenquellen verwaltet werden. Ebenso könnten die Datennutzenden Informationen über die Entstehung der sie betreffenden algorithmischen Ergebnisse bekommen, um diese besser einordnen zu können.

Mein *Prinzip der Maximierung der individuellen Kontrolle* soll die Pflicht, die Risikoexponierten so viel wie möglich über ihr Schadensrisiko selbst entscheiden zu lassen, verdeutlichen:

> Eine risikobehaftete Handlung ist nur dann zulässig, wenn den Risikoexponierten im Vorhinein so viel Kontrolle, wie unter den gegebenen Umständen möglich, zugestanden wurde. Eine Handlung ist hingegen nicht zulässig, wenn nicht alles Mögliche unternommen wird, um die Kontrolle der Risikoexponierten zu maximieren.

Diese moralische Pflicht begründe ich folgendermaßen:

(1) Wir haben eine Pflicht das eigene Handeln gegenüber anderen so zu gestalten, dass die anderen niemals bloß als Mittel, sondern immer auch als Selbstzweck behandelt werden (Kants dritte Formel des Kategorischen Imperativs).

(2) Wenn andere einem Risiko auf Schaden ausgesetzt werden, und ihnen dabei die Möglichkeit zur Kontrolle über das Risiko verwehrt wird, werden sie nicht als Selbstzweck behandelt und es ist möglich sie als (bloßes) Mittel zu gebrauchen.

(3) Um den Risikoexponierten die Möglichkeit zur Kontrolle über das Risiko nicht zu verwehren, muss man Risikoexponierten so viel wie möglich über ihr Schadensrisiko selbst entscheiden lassen.

→ Der Auferleger eines Risikos hat die Pflicht, die Risikoexponierten so viel wie möglich über ihr Schadensrisiko selbst entscheiden zu lassen.

Ganz konkret ist es dann zulässig, dass Datengenerierende und Datenverarbeitende im Datenökosystem Datenquellen und Datennutzenden Risiken auferlegen, wenn es innerhalb von Hanssons gerechtem sozialen System der Risikobereitschaft passiert, welches

den Risikoexponierten zum Vorteil gereicht.[523] Gilt die dritte Formel des Kategorischen Imperativs von Kant als eine der Regeln dieses Systems, entsteht eine deontologische Risikoethik. Diese besagt, dass Risikoexponierte über das Risiko informiert werden und sie idealerweise die Wahl haben sollten, ob sie das Risiko tragen wollen. Wenn dies nicht möglich ist, sollte das Risiko so gering wie möglich sein, und Risikoexponierte sollten für das Risiko sowie für mögliche Schäden Kompensation erhalten.[524] Zudem gilt bei der Gestaltung eines Datenökosystems, das Akteuren mit weniger Kontrolle Risiken auferlegt, mein Prinzip der Maximierung der individuellen Kontrolle. Dies soll sicherstellen, dass die Risikoexponierten im Rahmen des auferlegten Risikos immer noch Entscheidungsspielraum haben und dass sie besser verstehen können, worauf sie sich einlassen. Dieses Verständnis wiederum ist davon abhängig, dass relevante und verständliche Informationen zur Verfügung gestellt werden und dabei weder Zwang noch Täuschung vorliegen.

Im Folgenden soll ausführlich betrachtet werden, inwiefern die Datenquellen und Datennutzenden im Datenökosystem Risiken ausgesetzt werden und unter welchen Bedingungen dies Teil eines gerechten sozialen Systems der Risikobereitschaft sein kann. Denn bisher war nur relevant, ob das betreffende Risiko den Akteuren von außen auferlegt wird – diese Beschränkung der Autonomie ist nur unter den in diesem Kapitel herausgearbeiteten Bedingungen zulässig. Es stellt sich aber auch die Frage, warum die beiden unter 3.3 hergeleiteten Risiken überhaupt problematisch sein sollen. Denn entgegen weit verbreiteter Behauptungen sind die Konsequenzen einer Handlung für Kants Deontologie nicht unwichtig. Laut Hare sind aus deontologischer Sicht vielmehr genau die Handlungsfolgen relevant, welche die moralischen Grundsätze der Deontologie verbieten oder herbeiführen wollen.[525] Deshalb möchte ich im Folgenden das Argument anführen, dass das Verletzen der Privatsphäre und das Fehlen von Transparenz im Datenökosystem Handlungsfolgen sind, welche der Kategorische Imperativ verbietet. Es sollten generell keiner Person Risiken auferlegt werden, ohne dass sie die Wahl hat. Aber in den nächsten beiden Kapiteln wird begründet,

523 Hansson 2003:305
524 Ebd.:303
525 Hare 1993:15

warum es sich speziell bei der Verletzung der Privatsphäre und dem Fehlen der Transparenz um Risiken handelt, die vermieden werden sollten. Daraus werden dann administrative Anforderungen abgeleitet, welche den Schutz von Privatsphäre und Transparenz im Datenökosystem ermöglichen sollen, sodass die moralischen Grundsätze der Deontologie erfüllt werden können.

6 Privatsphäre

Die Verletzung der Privatsphäre ist ein Risiko, welchem die Datenquellen in einem Datenökosystem ausgesetzt werden. In diesem Kapitel wird erarbeitet, wieso die Verletzung der Privatsphäre überhaupt ein unerwünschtes Ereignis wäre, warum eine solche Verletzung eine verbotene Handlungsfolge wäre und wie man die Privatsphäre schützen bzw. fördern kann. Dafür ist es zuerst einmal notwendig zu definieren, was genau mit Privatsphäre gemeint ist – insbesondere wann Privatsphäre intakt ist und wann sie verletzt ist. Nur so können Bedingungen abgeleitet werden, unter denen es möglich ist, vor der Verletzung zu schützen und die intakte Privatsphäre zu fördern bei der Gestaltung eines Datenökosystems. Deshalb soll das Phänomen der Privatsphäre zunächst deskriptiv erfasst werden. Erst wenn klar definiert ist, unter welchen Umständen Privatsphäre intakt oder verletzt ist, soll ein normativer Maßstab angelegt werden und auch gefragt werden, ob es eine moralische Pflicht zum Schutz der Privatsphäre gibt. Wenn bei der rein deskriptiven Definition des Begriffes der Privatsphäre schon von einem Recht auf Privatsphäre die Rede ist, wird dieses Recht vorläufig als Anspruch verstanden, der sich aus einer moralischen Pflicht ableitet.

Der Politik- und Geschichtsphilosoph Raymond Geuss meint, öffentlich und privat lassen sich nicht klar voneinander unterscheiden, da sich diese beiden Sphären überschneiden. Seiner Meinung nach wird der Unterscheidung zwischen öffentlich und privat eine zu große Bedeutung zugeschrieben.[526] Es reiche nicht, auf die Frage, warum wir uns nicht einmischen sollten, „weil es privat ist“ zu antworten. Denn daraufhin müsste die Frage beantwortet werden, warum wir denken, dass wir uns in private Vorgänge nicht einmischen sollten. Geuss nimmt an, dass die Gründe sehr unterschiedlich wären und hier keine einheitliche Erklärung gefunden werden kann.[527]

526 Geuss 2001:5
527 Ebd.:107

Im Folgenden möchte ich zeigen, dass es doch eine einheitliche Erklärung gibt, und dass es bei dem Konzept der Privatsphäre weniger um die Ablehnung von Einmischung geht, sondern mehr um die Ablehnung von Kontrollverlust.

Laut Duden kommt das Wort „privat" vom lateinischen „privatus" und bedeutet „(der Herrschaft) beraubt; gesondert, für sich stehend; nicht öffentlich".[528] Der Jurist Thomas McIntyre Cooley orientiert sich bei seiner Definition von Privatsphäre sehr nah an dieser ursprünglichen Begriffsbedeutung: „The right to one's person may be said to be a right of complete immunity: to be let alone."[529] Ursprünglich wurde unter Privatsphäre also die Abgeschiedenheit von der Gesellschaft, das In-Ruhe-gelassen-werden verstanden.

Der Rechtsphilosoph und Ethiker William Parent ist der Meinung, dass die Definition von Privatsphäre als reine Abgeschiedenheit zu weit gefasst ist und dass sie den begrifflichen Kern der Privatsphäre völlig verkennt. Angenommen, eine Person A wird von B angegriffen. B hat A nicht in Ruhe gelassen, aber das bedeutet nicht, dass B in die Privatsphäre von A eingedrungen ist. Vielmehr sind zuerst einmal Begriffe wie Übergriff, Angriff und Gewalt hilfreich zum Beschreiben des Vorfalls. Das Konzept der Privatsphäre würde erst dann benötigt, wenn zum Beispiel B das Portemonnaie von A stiehlt und beim Ausräumen etwas Persönliches über A erfährt. Die Definition der Privatsphäre als Rückzug der Person von der Gesellschaft suggeriert fälschlicherweise, dass jedes Mal, wenn eine Person mit der Gesellschaft in Kontakt kommt, ihre Privatsphäre beeinträchtigt wird.[530] Hier wird deutlich, dass ein angemessenes Konzept der Privatsphäre nicht zulassen darf, dass die Privatsphäre einer Person bereits dadurch verletzt werden kann, indem sie in der Öffentlichkeit wahrgenommen wird.[531]

Es gibt verschiedene Dimensionen des Privaten: Rössler unterscheidet zwischen lokaler, dezisionaler und informationeller Privatheit.[532] Auch der Datenethiker und Technologiephilosoph Adam D.

528 Duden 2023

529 Cooley 1879:29; Oft werden als Autoren dieser Definition fälschlicherweise die Juristen Samuel D. Warren und Louis D. Brandeis genannt, welche sich jedoch auf Cooley beziehen (Warren, Brandeis 1890:195).

530 Parent 1983b:342

531 Ebd.:344

532 Rössler 2003:31

Moore macht einen Unterschied zwischen privaten Orten und Informationen – anstatt auf private Entscheidungen legt er aber den Fokus auf die Privatheit von Körpern.[533] Ich denke, die Privatheit von Körpern ist nicht zu vernachlässigen, da sie vor ungewollten Übergriffen schützen kann. Private Entscheidungen beziehen sich auf Entscheidungen, die über den eigenen Körper und private Orte hinaus gehen: „(...) mit wem will ich zusammenleben; welchen Beruf will ich ergreifen; aber auch: welche Kleidung trage ich (...)".[534] Auch hier ermöglicht die Privatsphäre einen Rückzug und dadurch eine Entscheidung ohne Beeinträchtigung. Die Dimensionen der Privatheit sind also Informationen, Orte, Körper und Entscheidungen.

Vor der Digitalisierung war die Zugangskontrolle zu Orten und Körpern das vorrangige Ziel der Privatsphäre, da alle persönlichen Informationen dort zu finden waren. Im Kontext der Digitalisierung sind persönliche Informationen allerdings von überall her zugänglich. Nun muss sich der Datenschutz an die technologischen Entwicklungen anpassen, indem er seinen Schwerpunkt auf den Schutz personenbezogener Daten verlagert. Die Wirtschaftswissenschaftler:innen Alessandro Acquisti, Laura Brandimarte und Jeff Hancock stimmen zu: „(…) the very boundaries of privacy have evolved with the digital transition—from predominantly physical to predominantly informational."[535] Da in dieser Arbeit der Umgang mit personenbezogenen Daten ethisch betrachtet wird, ist die Dimension der informationellen Privatheit besonders interessant. Im Folgenden ist deshalb mit Privatsphäre immer informationelle Privatsphäre gemeint.

Persönliche Informationen

Es fragt sich nun, was genau unter „persönliche Informationen" oder aber „personenbezogene Daten" fällt. Von dieser Definition ist abhängig, über welchen Gegenstand private Individuen die Kontrolle haben sollten. Die DSGVO bezeichnet als „(...) „personenbezogene Daten" alle Informationen, die sich auf eine identifizierte oder

533 Moore 2010:27
534 Rössler 2003:31
535 Acquisti et al. 2022:271

identifizierbare natürliche Person (...) beziehen (...)."[536] In gleicher Weise sind mit „persönlichen Informationen" laut O'Neill solche Informationen gemeint, welche ein Individuum identifizierbar machen: „The standard approach is to deem information 'personal' if it enables the identification of individuals."[537] Eine solche Definition wäre allerdings zu weit gefasst, da viele Informationen, die ein Individuum identifizierbar machen, öffentlich verfügbar sind, dadurch dass sich das Individuum zum Beispiel an öffentlichen Orten aufhält. So schreiben der Religionsphilosoph Neil Manson und die bereits erwähnte Ethikerin und politische Philosophin Onora O'Neill: „Personal information, in the sense that may be relevant to an account of informational privacy, cannot include publicly available knowledge about individuals."[538] Wenn es um frei verfügbare Informationen geht, von denen Außenstehende unweigerlich Kenntnis erlangen, kann sich das Individuum nicht auf die Privatsphäre berufen, um den Außenstehenden zu verbieten, von diesen Informationen Kenntnis zu erlangen.

Eine andere Möglichkeit wäre, nur die als persönliche Informationen zu definieren, die genau eine Person identifizierbar machen. Aber diese Definition ist wiederum zu eng, denn auch wenn die Information „Jemand hat Krebs" auf viele Individuen zutrifft, könnte jedes dieser Individuen ein Interesse daran haben, diese Tatsache über seine Gesundheit privat zu halten. Manson und O'Neill schlussfolgern: „So private information is at most a subset of the information that is true of a person and not already public knowledge."[539] Persönliche Informationen müssen also auf eine Person zutreffen und dürfen nicht bereits öffentlich bekannt sein. Um genauer zu definieren, was in diese Kategorie fällt, könnte man sich auf das Interesse an Privatsphäre fokussieren. Dadurch würde allerdings das Problem nur verlagert, da man spezifizieren müsste, was vernünftigerweise als intim oder sensibel anzusehen ist.[540]

Der Medizinethiker Andrea Martani und der Technologieethiker Patrik Hummel sind der Meinung, dass nicht nur personenbezoge-

536 Art. 4, Abs. 1, Satz 1 DSGVO
537 O'Neill 2021:81
538 Manson, O'Neill 2007:103
539 Ebd.
540 Ebd.:103f.

ne Daten ein Schutzgegenstand des Datenschutzes sein sollten, da Daten nicht personenbezogen sein müssen, um durch ihre Verarbeitung eine Gefahr für ein Individuum darzustellen.[541] Denn es ist möglich, dass aus nicht-persönlichen Informationen, persönliche Informationen über geschützte Merkmale abgeleitet werden.[542] Das geschieht durch sogenannte Inferenzen. Der Datenethiker Jakob Mainz erklärt ihre Funktionsweise mit Bezug auf Barocas und Nissenbaum folgendermaßen: „(…) pieces of information that are clearly personal can often be inferred from pieces of information that are clearly non-personal (…)."[543] Inferenzen sind Informationen, die durch Schlussfolgerungen oder Überlegungen und nicht durch direkte Beobachtung oder Erhebung bei der betroffenen Person gewonnen werden.[544] Durch solche indirekten Mittel ist es möglich, Rückschlüsse auf genau die Eigenschaften zu ziehen, die ohne identifizierende Informationen lange Zeit nicht bekannt zu sein schienen.[545]

Auf Grund eines Artikels von dem Journalisten Charles Duhigg im New York Times Magazine kursiert eine Anekdote, die ein gutes Beispiel liefert für solche möglichen Inferenzen: Eine junge Frau soll auf Grund ihres Kaufverhaltens bei Target (einer der größten Einzelhändler der USA) Gutscheine für Babykleidung und Kinderbetten zugeschickt bekommen haben. Sie soll wirklich schwanger gewesen sein, aber ihr Vater, bei dem sie wohnte, soll nichts davon gewusst haben und erst durch die zugesandten Gutscheine davon erfahren haben. Eine solche Schlussfolgerung über die Schwangerschaft einer Konsumentin lässt sich deshalb ziehen, weil andere Konsumentinnen preisgeben, dass sie schwanger sind. Diese schwangeren Frauen weisen ein bestimmtes Konsumverhalten auf und kauften zum Beispiel größere Mengen an unparfümierter Lotion sowie Nahrungsergänzungsmittel wie Kalzium, Magnesium und Zink. Wenn eine andere Frau das gleiche Konsumverhalten zeigt, wird daraus geschlossen, dass sie auch schwanger ist.[546] Die Anekdote beweist nicht ein-

541 Martani, Hummel 2022:37
542 Manson, O'Neill 2007:104; Wachter, Mittelstadt 2019:564; Martani, Hummel 2022:34
543 Mainz 2022:565, Fußnote 14; Vgl. Barocas, Nissenbaum 2014:55
544 Wachter und Mittelstadt 2019:515; Martani, Hummel 2022:33
545 Barocas, Nissenbaum 2014:55
546 Duhigg 2012; Hill 2012; Barocas, Nissenbaum 2014:62

deutig, dass Target die Schwangerschaft der jungen Frau aus ihrem Konsumverhalten abgeleitet hat[547], zeigt aber beispielhaft, welches Potenzial datenbasierte Inferenzen haben können.

Dieses Handeln auf Grundlage der Inferenzen hat dann aber nicht mehr notwendigerweise einen Bezug zur Privatsphäre des Einzelnen. Es wird zwar nach kleinen, aussagekräftigen Verbindungen zwischen Einzelpersonen gesucht, aber daraus werden Gruppenprofile erstellt, um noch mehr Informationen über Dritte herausfinden zu können.[548] Autoren wie der Technologiephilosoph und Digitalethiker Luciano Floridi behaupten, es handle sich hier um eine Gruppenprivatsphäre, die verletzt würde.[549] Er erklärt diese mit folgender Analogie: „The individual sardine may believe that the encircling net is trying to catch it. It is not. It is trying to catch the whole shoal. It is therefore the shoal that needs to be protected, if the sardine is to be saved."[550] Floridi argumentiert, dass bei der Datenverarbeitung eher Typen bzw. Gruppen behandelt werden als Einzelpersonen.[551] Denn Einzelpersonen werden in Gruppen eingeteilt von zum Beispiel schwangeren Frauen: „(…) most people are not targeted (…) as individuals but as members of specific groups (…)."[552] Laut Floridi können Einzelpersonen manchmal nur dadurch geschützt werden, dass die Gruppe geschützt wird, welcher die Einzelpersonen angehören.[553]

Die Gruppenprivatsphäre ist in solchen Kontexten relevant, in denen Inferenzen genutzt werden, um Nutzerpräferenzen, sensible Attribute und Meinungen abzuleiten oder Verhaltensweisen vorherzusagen.[554] So kann zum Beispiel von dem Online-Verhalten der Facebook Nutzenden politische Meinungen und der Gemütszustand abgeleitet werden und dann genutzt werden, um gezielt Werbung zu machen.[555] Auch im Kontext von Versicherungen oder Bepreisung von Produkten sollte die Gruppenprivatsphäre Beachtung finden, da

547 Fraser 2020
548 Wachter, Mittelstadt 2019:506
549 Floridi 2017
550 Floridi 2014:3
551 Ebd.:1f.
552 Ebd.:1
553 Ebd.:3
554 Wachter, Mittelstadt 2019:497
555 Merrill 2016; Reilly 2017; Wachter, Mittelstadt 2019:506f.

es hier möglich ist, dass die Menschen auf Grund der Zugehörigkeit zu einer Gruppe schlechter gestellt werden. In dieser Arbeit soll jedoch ausschließlich die individuelle Privatsphäre betrachtet werden, da es im Verwaltungskontext meist eine Person ist, die eine Verwaltungsleistung in Anspruch nimmt und dann individuelle Ergebnisse zur Verfügung gestellt bekommt. Selbst wenn eine Gruppe von Personen, wie zum Beispiel Eltern, gemeinsam eine Verwaltungsleistung in Anspruch nehmen, werden zur Bearbeitung des Antrags keine Gruppenprofile erstellt, sodass die Gruppenprivatsphäre keine Rolle spielt.

Trotzdem gibt die Möglichkeit der Inferenzen wichtige Hinweise für die Bestimmung von persönlichen Informationen, da hier potenzielle Schäden nicht dadurch entstehen, dass die verarbeiteten Daten personenbezogen waren, sondern dass sie verwendet wurden um Personen zu beeinflussen oder um Entscheidungen über sie zu fällen.[556] Die Rechtswissenschaftlerin Sandra Wachter und der bereits erwähnte Datenethiker Brent Mittelstadt fügen hinzu: „The belief that certain categories of data are fundamentally less harmful or risky than others is undermined by Big Data analytics."[557] Auf Grund der Möglichkeit, Inferenzen zu bilden, sei es schwierig, wirklich neutrale Informationen zu finden, die nicht in personenbezogene Daten umgewandelt werden können.[558] Dies zeigt auf, dass die Unterscheidung von persönlichen und nicht-persönlichen Informationen wenig hilfreich ist, da jegliche Information in Kombination mit anderen Informationen persönlich werden kann.

Auch Manson und O'Neill sind der Meinung, dass es nicht darum geht, welche Informationen persönlich sind und dementsprechend in den Bereich der Privatsphäre fallen. Denn Informationen an sich hätten keine ethische Bedeutung und seien nicht als solche schutzbedürftig: „What is informative and significant for one person is often meaningless or irrelevant to another."[559] Außerdem kann niemand den Anspruch erheben, dass Außenstehende bestimmte Informationen nicht über ihn/sie wissen dürfen.[560] Denn kein Individuum

556 Wachter, Mittelstadt 2019:616
557 Ebd.:616
558 Ebd.:518
559 Manson, O'Neill 2007:108
560 Thomson 1975:307

kann kontrollieren, was Außenstehende über es denken oder welche Schlüsse Außenstehende aus öffentlich verfügbaren Informationen über das Individuum ziehen.[561]

Manson und O'Neill machen stattdessen auf die Art der Beschaffung der Daten aufmerksam: „The acquisition of 'personal' information is often an unavoidable or highly likely consequence of morally permissible action. However, in some cases the routes or means by which such knowledge is acquired may be intrusive and impermissible."[562] Deshalb geht es bei der Privatsphäre nicht um den Inhalt der Informationen, sondern um Handlungen und Kommunikation, die zur Beschaffung von „persönlichen" Informationen führen können, sowie um die Verwendungen solcher Informationen. Mainz zufolge sind Inferenzen zulässig, solange das Erwerben der Daten zulässig war. Eine Inferenz verletzt jedoch die Privatsphäre der Betroffenen, wenn die für die Inferenz notwendigen Informationen nicht rechtmäßig erlangt wurden.[563]

Auch laut Manson und O'Neill muss unterschieden werden zwischen moralisch zulässigen und moralisch unzulässigen Handlungen, welche zur Beschaffung und Verarbeitung von Informationen über andere führen.[564] Solche Handlungen können moralisch unzulässig sein, wenn sie zum Beispiel unverständlich, irrelevant, zwanghaft, schädlich, belastend, irreführend oder unehrlich sind, und damit gegen ethische und epistemische Normen verstoßen.[565] Deshalb sollte es für diejenigen, die Informationen über andere beschaffen und verarbeiten, bestimmte Pflichten geben.[566] In dieser Arbeit wurde Kants dritte Formel des Kategorischen Imperativs als eine solche Pflicht für das Sammeln und Verarbeitung von Daten eingeführt – das heißt, niemand darf dabei als bloßes Mittel gebraucht werden. Bei der informationellen Privatsphäre geht es also weniger darum, welche Informationen privat sind, sondern mehr darum, wie die

561 Manson, O'Neill 2007:106
562 Ebd.
563 Mainz 2022:569
564 Ebd.:110
565 Ebd.:121
566 Ebd.:111

Informationen erlangt werden und zu welchem Zweck sie verarbeitet werden.[567]

6.1 Kriterien für eine Definition

Es stellt sich weiterhin die Frage, wann Privatsphäre intakt und wann sie verletzt ist. Hierzu gibt es verschiedene Auffassungen in der Literatur, wie zum Beispiel das Verständnis von Privatsphäre als das Recht in Ruhe gelassen zu werden[568] – auch der Zugangs-Ansatz und der Kontroll-Ansatz werden im Folgenden eine Rolle spielen. Aber zuerst muss geklärt werden, wie festgestellt werden kann, ob sich ein Ansatz als Definition von Privatsphäre eignet.

Der Ethiker, Sozial- und politische Philosoph Leonhard Menges identifiziert in der Literatur ein Standardverfahren zur Prüfung der Angemessenheit von Definitionen der Privatsphäre. Dafür wird eine beispielhafte Situation entwickelt, in der es intuitiv plausibel ist, dass ein Akteur in Bezug auf eine bestimmte Information Privatsphäre hat. Auch die zu prüfende Definition von Privatsphäre sollte in dieser Situation zu dem Schluss kommen, dass die Privatsphäre des betroffenen Akteurs gegeben ist. Als nächstes stellt man sich ein Ereignis vor, das die Situation in einer Weise verändert, die entweder (a) intuitiv die Privatsphäre der Person nicht beeinträchtigt oder (b) intuitiv die Privatsphäre der Person beeinträchtigt. Die Definition von Privatsphäre wird dann geprüft, indem sie auf die durch das Ereignis veränderte Situation angewendet wird. Wenn das Ereignis (a) intuitiv die Privatsphäre der Person nicht beeinträchtigt, die Definition aber impliziert, dass dies der Fall ist, dann ist die Definition zu weit gefasst und impliziert falsch-positive Ergebnisse. Wenn das Ereignis (b) intuitiv die Privatsphäre der Person beeinträchtigt, die

567 Im Folgenden wird angenommen, dass die Daten im Verwaltungskontext immer Individuen betreffen und es deshalb relevant ist, wie die Informationen erlangt werden und zu welchem Zweck sie verarbeitet werden. Die Begriffe ‚persönlich' und ‚personenbezogen' werden im Folgenden in Bezug auf Informationen und Daten verwendet, um zu kennzeichnen, dass Individuen betroffen sind.

568 Cooley 1879:29; Warren, Brandeis 1890:195

Definition aber impliziert, dass dies nicht der Fall ist, dann ist die Definition zu eng und impliziert falsch-negative Ergebnisse.[569]

Beispiele für Ereignisse vom Typ (a), durch welche die Privatsphäre intuitiv nicht beeinträchtigt wird, sind sogenannte *Threatened Loss Cases*. Ein Beispiel dieser Art von der Moralphilosophin und Metaphysikerin Judith Jarvis Thomson, weiterentwickelt von Parent und Menges, ist ein Streit zwischen zwei Partner:innen in ihrem eigenen Haus – es sei intuitiv plausibel, dass die streitenden Personen in Bezug auf die Information, dass sie streiten und worüber sie streiten, Privatsphäre haben. Das Ereignis, durch welches die Lage der Privatsphäre neu bewertet werden muss, ist die Erfindung eines Röntgengeräts, das durch Wände schauen kann und auf das Haus der Streitenden gerichtet ist, aber nicht benutzt wird – intuitiv wird dadurch laut Thomson die Privatsphäre der Personen nicht beeinträchtigt[570]: „(...) my right to privacy is not violated until my neighbor actually does train the device on the wall of my house. It is the actual looking that violates it, not the acquisition of power to look."[571] Ein weiteres Beispiel dieser Art von dem Technologieethiker Kevin Macnish ist eine Situation, in der eine Person im Café persönliche Gefühle und Gedanken in ihr Tagebuch schreibt – es sei intuitiv plausibel, dass die Person in Bezug auf diese Gefühle und Gedanken Privatsphäre hat. Dann verlässt aber die Person das Café und vergisst ihr Tagebuch auf dem Tisch. Eine andere Cafébesucherin legt das Tagebuch auf ihren Tisch, ohne es zu lesen und wartet auf die Rückkehr der Person – intuitiv werde dadurch die Privatsphäre der Person nicht beeinträchtigt.[572]

Beispiele für Ereignisse vom Typ (b), durch welche die Privatsphäre intuitiv beeinträchtigt wird, sind *Voluntary Divulgence Cases*. Ein Beispiel dieser Art ist eine Situation, in der jemand wegen Selbstzweifeln eine Flasche Schnaps am Tag trinkt, aber niemand davon weiß – es sei intuitiv plausibel, dass die Person in Bezug auf diese Information Privatsphäre hat. Das Ereignis, durch welches die Lage der Privatsphäre neu bewertet werden muss, ist, dass die Person freiwillig und nüchtern fremden Menschen von ihren persönli-

569 Menges 2022:697
570 Parent 1983b:344; Menges 2022:697
571 Thomson 1975:305, Fußnote 1
572 Macnish 2018:420

chen Problemen und dem regelmäßigen Alkoholkonsum erzählt – intuitiv werde dadurch die Privatsphäre der Person beeinträchtigt.[573]

Das Problem an diesem Standardverfahren zur Prüfung der Angemessenheit von Definitionen der Privatsphäre ist, dass womöglich nicht alle Menschen die gleiche Intuition haben bezüglich dieser Beispiele.[574] Wenn wir aber die Definition von Privatsphäre in verschiedenen Beispielszenarien ausschließlich an Intuitionen messen, dann müssten alle die gleiche Intuition haben, damit diese Messungen allgemeingültig sind. Zu folgendem Beispiel gibt es sogar in der Literatur unterschiedliche Intuitionen: Der Wind weht das Tagebuch aus dem offenen Rucksack einer Person, und es landet mit den Seiten nach oben auf dem Bürgersteig. Eine andere Fußgängerin ist so freundlich, das Tagebuch aufzuheben, doch dabei kann sie nicht umhin, einen Teil des Inhalts zu lesen. Die Datenethikerin Haleh Asgarinia vertritt die Intuition, dass in diesem Beispiel die Privatsphäre der Tagebuchbesitzerin verletzt wird, da die Fußgängerin einen Teil des Tagebuchinhalts liest, obwohl die Tagebuchbesitzerin sich nicht wünscht, dass die Fußgängerin das Tagebuch liest.[575] Den Datenethikern Jakob Thrane Mainz und Rasmus Uhrenfeldt zufolge wird dadurch jedoch intuitiv die Privatsphäre der Person nicht beeinträchtigt, weil der Kontrollverlust über das Tagebuch nicht auf die Handlung eines anderen Akteurs zurückzuführen ist.[576] Hier zeigt sich, dass sich die Debatte zum Teil darum dreht, welche Intuitionen richtig sind.[577] Insbesondere, wenn es gegensätzliche Intuitionen in Bezug auf eine Situation gibt, wird deutlich, dass es problematisch ist, eine Intuition als Kriterium für die Definition eines Begriffes zu betrachten.

Deshalb wird in dieser Arbeit die Angemessenheit von Definitionen der Privatsphäre nicht anhand von Gegenbeispielen geprüft. Stattdessen lässt sich allein auf Grund der Fragestellung in diesem Kapitel ein Kriterium für eine Definition von Privatsphäre aufstellen: Eine Definition der Privatsphäre sollte die Möglichkeit bieten, nicht nur zu bestimmen, wann die Privatsphäre durch Außenstehende verletzt wird, sondern auch wann sie intakt ist.

---

573 Parent 1983a:273; Menges 2022:698
574 Menges 2020c:394; Lundgren 2021a:672
575 Asgarinia 2023:335
576 Mainz, Uhrenfeldt 2021:298
577 Lundgren 2021a:672

6.1.1 Individuelle Intuitionen

Die unterschiedlichen Intuitionen bezüglich der Bedeutung von Privatsphäre lassen sich unter anderem auf unterschiedliche kulturelle Prägungen zurückführen. Denn es lässt sich beobachten, dass Privatsphäre in verschiedenen Kulturen unterschiedlich gelebt wird. Der Jurist Charles Fried stimmt zu: „(...) the concrete expressions of privacy in particular societies and cultures differ enormously."[578] So hängt es zum Beispiel von der Kultur ab, ob eine Berührung eine Verletzung der Privatsphäre darstellt. Auch laut Moore gibt es je nach Kultur unterschiedliche normative Ansprüche in Bezug auf die Privatsphäre.[579] Man könnte auch vermuten, dass verschiedene Generationen der gleichen Kultur ihre Privatsphäre unterschiedlich ausleben. Insbesondere durch die Entstehung von Sozialen Medien ist es seit ca. 30 Jahren möglich, Informationen über sich mit vielen fremden Menschen zu teilen. Es sind vor allem jüngere Menschen – die Generationen, die mit Sozialen Medien aufgewachsen sind –, welche diese Möglichkeit nutzen und persönliche Informationen online teilen. Solche Informationen, zum Beispiel über ihre sexuelle Orientierung oder ihre politische Einstellung, teilen manche Menschen aus älteren Generationen der gleichen Kultur nicht einmal mit ihrem engsten Umfeld.

Ich argumentiere jedoch, dass das Ausleben der Privatsphäre sogar noch individueller ist. Denn jede Person kann individuell entscheiden, ob sie Informationen teilt, welche und mit wem. Deshalb gibt es in jeder Kultur und in jeder Generation Menschen, die sich aktiv dafür entscheiden, persönliche Informationen in Sozialen Medien zu teilen. Der Zweck kann Selbstdarstellung sein, das Finden einer Gruppe von Gleichgesinnten oder auch das Verdienen seines Lebensunterhalts. Die Gruppe von Nutzenden, die Zugriff auf die Informationen bekommen, kann beschränkt werden oder bewusst nicht eingeschränkt werden. Außerdem werden ganz bestimmte Informationen geteilt, viele andere persönliche Informationen behalten die auf Sozialen Medien Aktiven aber für sich. Gleichzeitig gibt es in jeder Kultur und in jeder Generation Menschen, die ihre per-

578 Fried 1968:213
579 Moore 2010:26

sönlichen Informationen nicht online teilen – das Gleiche gilt im analogen Austausch mit Mitmenschen.

Fried zufolge kann in konkreten Situationen und Gesellschaften die Möglichkeit, frei zu entscheiden, ob man Informationen über die eigene Person teilt, relativ und eingeschränkt sein.[580] Denn wenn in einer Gesellschaft der Zugang zu Sozialen Medien beschränkt ist oder das Veröffentlichen von Informationen über zum Beispiel die sexuelle Orientierung oder die politische Einstellung negative Konsequenzen nach sich ziehen könnte, dann ist es nicht möglich bzw. nicht ratsam bestimmte Informationen auf diesem Weg zu teilen. Das heißt, die Möglichkeit, die eigene Privatsphäre individuell zu gestalten, kann von außen beschränkt sein. Trotzdem können wir hierdurch etwas über die Privatsphäre lernen. So schreibt Moore: „The forms of privacy may be culturally relative, but the need for privacy is not."[581] Unabhängig davon, wie eine Person ihre Privatsphäre versteht und auslebt, braucht sie Privatsphäre.[582]

Dadurch wird klar, dass die Definition von Privatsphäre individuellen Intuitionen gerecht werden muss. Insbesondere weil das individuelle Empfinden darüber, welche Informationen persönlich sind, variieren kann, sollte die individuelle Entscheidung darüber, welche Informationen geteilt werden, respektiert werden. Deshalb ist ein weiteres Kriterium für eine Definition von Privatsphäre, neben der Bestimmung von intakter und verletzter Privatsphäre, dass Individuen ihre Privatsphäre unterschiedlich empfinden und ausleben können müssen.

6.2 Ansätze zur Definition von Privatsphäre

Um einer Definition von Privatsphäre näher zu kommen, welche Raum lässt für individuelle Intuitionen und es ermöglicht zu bestimmen, wann Privatsphäre intakt und wann verletzt ist, werden im Folgenden zwei in der Literatur prominente Ansätze vorgestellt und

580 Fried 1968:213
581 Moore 2010:56
582 Warum wird erklärt in Kapitel 6.3

anhand dieser Kriterien bewertet: der Zugangs-Ansatz und der Kontroll-Ansatz.[583]

6.2.1 Zugangs-Ansatz

Der bereits erwähnte Technologieethiker Kevin Macnish argumentiert, dass ein Verlust der Privatsphäre nur dann eintritt, wenn auf die eigenen Daten tatsächlich zugegriffen wird. Aus Sicht des Zugangs-Ansatzes stellt das Sammeln der Informationen allein keinen Eingriff in die Privatsphäre dar – erst wenn tatsächlich auf die Informationen zugegriffen wird, wird die Privatsphäre verletzt.[584] Zur Illustration schlägt Macnish das unter 6.1 erwähnte Tagebuchbeispiel vor: Eine Person vergisst ihr Tagebuch auf dem Tisch im Café, und als sie nach 30 Minuten zurückkommt, liegt das Tagebuch auf dem Tisch einer Fremden. Laut Macnish gehen starke Gefühle mit einem solchen Kontrollverlust einher, aber seiner Meinung nach ist es ein Fehler, diese Gefühle mit einer Verletzung der Privatsphäre in Verbindung zu bringen. Wenn die Person ihr Tagebuch auf dem Tisch einer Fremden sieht, befürchtet sie, dass ihre Privatsphäre verletzt wurde, und das ist möglich. Allerdings wurde laut Macnish nur dann die Privatsphäre verletzt, wenn das Tagebuch tatsächlich geöffnet und gelesen wurde. In diesem Beispiel hat die fremde Café Besucherin das Tagebuch auf ihren Tisch gelegt, ohne es zu lesen und hat aufgepasst, dass auch niemand anderes im Tagebuch liest. Das bedeutet aus der Sicht von Macnish, dass die Privatsphäre der Besitzerin des Tagebuchs nicht verletzt wurde.

583 Es gibt auch andere, viel beachtete Ansätze zur Definition von Privatsphäre, die hier nicht diskutiert werden, wie zum Beispiel den Ansatz der „contextual integrity“ von Nissenbaum. Ihrer Meinung nach ist die Privatsphäre intakt, solange der Informationsfluss mit den kontextspezifischen Informationsnormen übereinstimmt (Nissenbaum 2010:188; Barocas, Nissenbaum 2014:47). Allerdings erlauben die für alle gleichermaßen geltenden Informationsnormen keinen Raum für individuelle Intuitionen. So wird Nissenbaums Ansatz nicht den Tatsachen gerecht, dass es viele verschiedene Interpretationen dessen gibt, wann Informationen mit dem Kontext konsistent verwendet werden, und dass unterschiedliche Individuen unterschiedliche Verwendungen ihrer Daten bevorzugen (Rule 2019:274).

584 Macnish 2018:417f.

Macnish trennt Verletzung der Privatsphäre von anderen Schäden wie zum Beispiel der Verringerung der Sicherheit: „(...) there are other harms associated with privacy violations besides the loss of privacy."[585] In einer anderen Version des Tagebuchbeispiels hat die Fremde im Café das Tagebuch gelesen und erpresst die Person, der das Tagebuch gehört, indem sie mit einer Veröffentlichung der Inhalte droht. In diesem Fall wurde, auch nach Macnish's Verständnis, die Privatsphäre der Person verletzt, aber es gibt einen weiteren Schaden: Die Verringerung der Sicherheit der Person auf Grund ihrer Erpressung. In einer weiteren Version des Tagebuchbeispiels gibt die Fremde im Café nur vor, das Tagebuch gelesen zu haben, und versucht die Person zu erpressen, indem sie mit Veröffentlichung der Informationen droht. Hier würde Macnish behaupten, dass die Privatsphäre der Person unverletzt bleibt, aber der weitere Schaden an der Sicherheit der Person bleibt bestehen: „(...) it is the reduction in security that is the greater harm."[586] Er hält die Verletzung der Sicherheit sogar für größer als die (potenzielle) Verletzung der Privatsphäre.

Der Technologieethiker Björn Lundgren ist wie Macnish ein Vertreter des Zugangs-Ansatzes[587] und meint, dass der Anspruch auf Privatsphäre auch einen Schutz vor bestimmten Risiken für die Privatsphäre umfasst: „(...) the right to privacy includes a right not to have others put one's privacy at substantial risk."[588] Da wir Menschen uns in einer Gesellschaft aber täglich gegenseitig Risiken aussetzen, kann nicht jedes Risiko für die Privatsphäre verboten sein. Deshalb kann dieser Anspruch auf Privatsphäre und Nicht-Risikoexposition auch außer Kraft gesetzt werden, wenn es zum Beispiel unvermeidbar oder notwendig ist, persönliche Informationen von anderen einzuholen.[589] Verletzt wird die Privatsphäre von Individuen dann, wenn durch Außenstehende ihre Überzeugungen oder ihr Wissen über den Zugang zu ihren privaten Angelegenheiten erheblich beeinträchtigt werden.[590]

585 Macnish 2018:425
586 Ebd.:425
587 Lundgren 2020:172f.
588 Lundgren 2021b:385f.
589 Ebd.:384–387
590 Ebd.:390

Das heißt, laut Lundgren verletzen Außenstehende die Privatsphäre eines Individuums dann,

- wenn sie tatsächlich auf persönliche Informationen des Individuums zugreifen[591],
- wenn sie das Individuum einem substanziell größeren Risiko aussetzen, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift[592],
- wenn sie dafür sorgen, dass das Individuum glaubt, dass seine privaten Angelegenheiten dem Risiko des Zugriffs durch eine andere außenstehende Person ausgesetzt sind[593] oder
- wenn sie dafür sorgen, dass das Individuum unsicher ist, ob seine privaten Angelegenheiten dem Risiko des Zugriffs durch eine andere außenstehende Person ausgesetzt sind.[594]

An dieser Stelle stellt sich die Frage, wann genau es zu einem tatsächlichen Zugang zu persönlichen Daten kommt. Mainz und Uhrenfeldt sind der Meinung, dass ein tatsächlicher Zugang dann möglich ist, wenn eine außenstehende Person eine epistemische Beziehung zu der betreffenden Information aufgebaut hat.[595] Macnish postuliert, dass die Privatsphäre nur dann beeinträchtigt ist, wenn diejenigen, die auf persönliche Informationen zugreifen, die Informationen verstehen können.[596] Das heißt, wenn die Informationen im Tagebuch für die Fremde im Café gar nicht verständlich sind, weil sie zum Beispiel die Sprache nicht spricht, in welcher das Tagebuch verfasst ist, dann wird keine epistemische Beziehung zu der betreffenden Information aufgebaut und die Privatsphäre des Tagebuchbesitzers bleibt intakt.[597] Die Datenethiker Lauritz Munch und Jakob Mainz sind aber der Meinung, dass für eine Verletzung der Privatsphäre neben der epistemisch begründeten Überzeugung in Bezug auf die persönlichen Informationen explizit auch eine Wahrnehmung der Information erfolgen muss[598]: „If no one forms

591 Macnish 2018:417f.; Lundgren 2021b:387
592 Lundgren 2021b:387
593 Ebd.:389
594 Ebd.
595 Mainz, Uhrenfeldt 2021:291f.
596 Macnish 2020; Asgarinia 2023:338
597 Munch, Mainz 2023:252f.
598 Ebd.:247f.

a warranted belief about A's personal matters, and no one perceives them, then A has full privacy."[599] Die Autoren entwickeln folgendes Beispiel, um die Relevanz der Wahrnehmung einer Information zu unterstreichen: B benutzt ein Röntgengerät, um sich den Inhalt des Safes von A anzuschauen. Dadurch sieht B ein Nacktfoto von A, das in dem Safe aufbewahrt wird. B glaubt jedoch, dass er halluziniert und meint fälschlicherweise, dass er das Foto nicht richtig wahrnimmt.[600] Laut Munch und Mainz wird hier die Privatsphäre verletzt, weil B den Inhalt von A's Safe wahrnimmt – unabhängig davon, dass B als Reaktion auf diese Wahrnehmungserfahrung weder eine Überzeugung bildet noch eine Rechtfertigung aus ihr gewinnt.[601]

6.2.1.1 Argumente gegen den Zugangs-Ansatz

Munch und Mainz erklären, dass sie sich neben der Bildung begründeter Überzeugungen und der tatsächlichen Wahrnehmung keine weiteren Prozesse vorstellen können, welche die Privatsphäre auf plausible Weise beeinträchtigen.[602] Ich denke jedoch, ihr eigenes Beispiel zeigt eindrücklich, dass allein der Versuch, eine epistemische Beziehung zu der betreffenden Information aufzubauen, zu der einem nicht aktiv Zugang gewährt wurde, eine Verletzung der Privatsphäre darstellt. Denn ich denke, das Problem ist, dass B in den Safe von A schaut. Es geht nicht darum, was B wahrnimmt und woran er sich erinnert, sondern das B überhaupt in den Safe von A schaut. Das Verhalten von B, sich Zugang zu verschaffen zu Informationen, die A offensichtlich nicht mit ihm teilen möchte, ist das Problem – das Umgehen von As Kontrolle über diese Informationen ist das, was ihre Privatsphäre verletzt. Ob B den Inhalt vom Safe wahrnimmt und sich auch darüber bewusst ist, dass er ihn wahrnimmt, ist unwichtig – selbst wenn der Safe leer wäre, hätte B durch diesen unerwünschten Zugriff As Privatsphäre verletzt.

Einerseits werden also durch den Zugangs-Ansatz *unerwünschte* Zugriffe nicht früh genug als Verletzung der Privatsphäre eingestuft

599 Munch, Mainz 2023:255
600 Ebd.:249
601 Ebd.:253
602 Ebd.:256

– es muss erst eine Wahrnehmungserfahrung geben[603], eine epistemische Beziehung zur Information aufgebaut sein[604], oder die Information verstanden werden[605]. Andererseits werden durch den Zugangs-Ansatz *erwünschte* Zugriffe zu früh als Verletzung der Privatsphäre eingestuft. So schreibt Asgarinia: „(…) this view implies that we lose our privacy when we freely and knowingly tell our friends about our problems and secrets."[606] Nach dem Zugangs-Ansatz wird unsere Privatsphäre immer dann beeinträchtigt, wenn ein anderer auf unsere persönlichen Daten zugreift, unabhängig davon, ob wir die Absicht haben, unsere persönlichen Daten mit einem anderen Akteur zu teilen. So wird es also auch als eine Verletzung der Privatsphäre eingestuft, wenn der Zugang durch das Individuum gewährt wurde. Aber laut der Philosophin Julie Inness und laut Menges können wir uns bewusst anderen Menschen mitteilen und andere in unseren Bereich der Privatsphäre einbeziehen, ohne unsere Privatsphäre zu beeinträchtigen.[607]

Nicht nur die Bestimmung dessen, wann die Privatsphäre verletzt ist, stößt auf Probleme. Auch die Bestimmung dessen, wann die Privatsphäre intakt ist, löst Schwierigkeiten aus. Laut Munch und Mainz ist die Privatsphäre einer Person intakt, wenn sich niemand eine begründete Meinung über ihre persönlichen Angelegenheiten bildet und niemand diese wahrnimmt.[608] Das heißt, die Privatsphäre ist intakt, wenn niemand Zugang zu den Informationen des Individuums hat und dementsprechend Außenstehende nicht auf Informationen über das Individuum zugreifen. Wie in Tabelle 1 deutlich wird, erinnert diese Definition von Privatsphäre an das Recht in Ruhe gelassen zu werden aus Teilkapitel 6.1.[609] Das Problem bei dieser Definition von Privatsphäre als reine Abgeschiedenheit ist, dass sie zu weit gefasst ist, denn sie suggeriert, dass jedes Mal, wenn eine Person mit der Gesellschaft in Kontakt kommt, ihre Privatsphäre beeinträchtigt wird.[610]

---

603 Munch, Mainz 2023:253
604 Mainz, Uhrenfeldt 2021:291f.
605 Munch, Mainz 2023:252f.
606 Asgarinia 2023:338
607 Inness 1992:46; Menges 2022:702f.
608 Munch, Mainz 2023:255
609 Cooley 1879:29; Warren, Brandeis 1890:195
610 Parent 1983b:342

Tabelle 1 Teile der Privatsphäre nach dem Zugangsansatz[611]

Zugangs-Ansatz	Individuum	Außenstehende
Privatsphäre intakt	Niemand bildet sich eine begründete Meinung über die persönlichen Angelegenheiten des Individuums und niemand nimmt sie wahr.[612]*	Außenstehende haben keinen Zugang zu den Informationen des Individuums, bilden sich keine begründete Meinung über seine persönlichen Angelegenheiten und nehmen diese auch nicht wahr.[613]*
Privatsphäre verletzt	Jemand hat Zugang zu den Informationen des Individuums, bildet sich eine begründete Meinung über die persönlichen Angelegenheiten des Individuums und nimmt sie wahr.[614]*	Außenstehende greifen tatsächlich auf persönliche Informationen des Individuums zu.[615] ODER Außenstehende sorgen dafür, dass das Individuum einem substanziell größeren Risiko ausgesetzt ist, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift.[616] ODER Außenstehende sorgen dafür, dass das Individuum glaubt, dass ODER unsicher ist, ob seine privaten Angelegenheiten dem Risiko des Zugriffs durch eine andere außenstehende Person ausgesetzt sind.[617]
* Erinnert an das Recht in Ruhe gelassen zu werden.[618]		

Das führt zu Schwierigkeiten mit den von mir aufgestellten Kriterien für eine Definition von Privatsphäre. Denn mit Hilfe des Zugangs-Ansatzes werden *unerwünschte* Zugriffe nicht früh genug und *erwünschte* Zugriffe zu früh als Verletzung der Privatsphäre eingestuft. Auch die Bestimmung, wann die Privatsphäre intakt ist, führt den Zugangs-Ansatz zu der Annahme, dass jede Wahrnehmungserfahrung eine Gefahr darstellt – dann wäre es beim gesellschaftlichen

611 Eigene Darstellung
612 Munch, Mainz 2023:255
613 Ebd.
614 Ebd.
615 Macnish 2018:417f.; Lundgren 2021b:387
616 Lundgren 2021b:387
617 Ebd.:389
618 Cooley 1879:29; Warren, Brandeis 1890:195

Zusammenleben jedoch sehr schwer eine intakte Privatsphäre zu erhalten. Insbesondere durch die Einstufung erwünschter Zugriffe als Verletzungen der Privatsphäre bietet der Zugangs-Ansatz keinen Raum für individuelle Intuitionen und das Ausleben der Privatsphäre nach den eigenen Vorstellungen.

Um das letztere Problem zu lösen schlägt Asgarinia eine Revision des Zugangs-Ansatzes vor: „A's privacy is lost iff: B actually accesses personal information P about A, and A intends that P remain private, or, B has reason to think that A intends to keep it private."[619] Nur wenn eine Person ihre persönlichen Informationen gerne für sich behalten würde, stellt der Zugang zu diesen Informationen eine Verletzung ihrer Privatsphäre dar. Diese Intention, die Informationen für sich zu behalten, kann entweder konkret geäußert werden, oder die Außenstehenden können diese antizipieren. Bei einem Tagebuch kann man sich zum Beispiel denken, dass der gesamte Inhalt privat bleiben soll. Im Allgemeinen gibt es viele Situationen, in denen der Zugang zu Informationen ausschließlich zu einem bestimmten Zweck gewährt wird. So zum Beispiel, wenn man seine Bankdaten bei einem Online-Einkauf angibt – da sollen die Informationen nur dazu benutzt werden, um den Einkauf abzuwickeln, aber nicht für andere Zwecke. Diese Form des Zugangs-Ansatzes plädiert also für zweckgebundene Verarbeitung von persönlichen Informationen – immer mit Rücksicht auf die Intention des Individuums.

Asgarinias Revision schärft die Bestimmung dessen, wann die Privatsphäre verletzt wird – nämlich dann, wenn jemand tatsächlich Zugang zu Informationen über eine Person hat, obwohl die Person die Information geheim halten wollte. Mainz und Uhrenfeld kritisieren Asgarinias Revision, denn sie meinen die Unerwünschtheit des Zugangs könne keine notwendige Bedingung sein für die Verletzung der Privatsphäre, da es Fälle geben kann, in denen die Privatsphäre durch einen Zugang verletzt wird, der durch die verletzte Person gewollt ist.[620] Man kann eine Privatsphärenverletzung wollen, wenn man sich zum Beispiel heimlich wünscht, dass man belauscht wird, während man eine Entschuldigung ausspricht, damit man die Ent-

619 Asgarinia 2023:340; Kursive Betonung entfernt

620 Man kann eine Privatsphärenverletzung wollen, wenn man sich zum Beispiel heimlich wünscht, dass man belauscht wird, während man eine Entschuldigung ausspricht, damit man die Entschuldigung nicht nochmal vortragen muss (Mainz, Uhrenfeldt 2021:297).

schuldigung nicht nochmal vortragen muss. Laut Mainz und Uhrenfeldt bleibt das Belauschen aber trotz Wunsch der sprechenden Person eine Verletzung der Privatsphäre. Das heißt, die Privatsphäre kann auch verletzt werden, wenn jemand tatsächlich Zugang zu Informationen über eine Person hat und die Person sich das heimlich gewünscht hat.

Asgarinias Revision gibt trotzdem einen interessanten Hinweis, da sie im Umkehrschluss klarer macht, wann die Privatsphäre intakt ist; nämlich dann, wenn niemand Zugang hat oder die Person ihre Informationen nicht geheim halten will und deshalb Zugang gewährt. Damit bietet Asgarinias Revision Raum für individuelle Intuitionen, in Form von zum Beispiel dem Bedürfnis guten Freunden Zugriff auf die eigenen Informationen zu gewähren. Diese Möglichkeit aktiv Zugang zu gewähren, erinnert an den Kontroll-Ansatz, der im Folgenden vorgestellt wird.

6.2.2 Kontroll-Ansatz

Beim Kontroll-Ansatz stellt die selektive Offenlegung den Kern der Privatsphäre dar.[621] Viele Autor:innen definieren Privatsphäre in dem Zuge als Möglichkeit zur Kontrolle.[622] Laut Fried ist Privatsphäre die Kontrolle über das Wissen über die eigene Person. Das heißt, eine Person, die Privatsphäre genießt, ist in der Lage, anderen den Zugang zu Wissen über sich selbst zu gewähren oder zu verweigern. Dabei gelte nicht, dass wir umso mehr Privatsphäre haben, je weniger über uns bekannt ist: „Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves."[623] Hierbei gehe es nicht nur um die Kontrolle der Quantität der Informationen, sondern auch um die Qualität des Wissens: „We may not mind that a person knows a general fact about us, and yet feel our privacy invaded if he knows the details."[624] Auch Rössler definiert etwas als privat, wenn man selbst den Zugriff auf dieses „Etwas" kontrollieren kann: „(…)

621 Parent 1983b:343
622 Westin 1967; Fried 1968, Rachels 1975, Inness 1992, Rössler 2004; Moore 2010
623 Fried 1970:140; Fried 1968:209
624 Fried 1968:210

the protection of privacy means protection against unwanted access by other people."[625] Ist die Privatsphäre einer Person geschützt, so ist die Person geschützt vor einem unerwünschten Zugriff durch andere Personen. Es handelt sich hierbei um den Schutz vor Einspruchs- oder Rückschlussmöglichkeiten anderer, wenn es um bestimmte private Entscheidungen, Handlungen, Verhaltensweisen oder Lebensweisen geht.[626]

Moore definiert das Recht auf Privatsphäre als Recht zur Kontrolle über Zugang zu und Benutzung von persönlichen Informationen.[627] Er meint, die Privatsphäre ist dann intakt, wenn eine Person sich freiwillig von ihren Mitmenschen abgrenzt und den Zugang einschränkt. Selbst wenn der Zugang möglich ist, folgt daraus nicht zwangsläufig, dass alle nachfolgenden Verwendungen der durch diesen Zugang erlangten Informationen ebenfalls zulässig sind.[628] Deshalb legt Moore Wert darauf, dass Privatsphäre nicht nur Kontrolle über den Zugang zu persönlichen Informationen ermöglicht, sondern auch Kontrolle über die Benutzung der zugänglich gemachten Informationen. Ein Fallbeispiel macht den Effekt der ergänzten Benutzung von Informationen deutlich: Jones hört zufällig mit, als Smith seinem Freund erzählt, dass er geschieden ist, und spricht mit anderen Freunden darüber.[629] Jones bekommt zufällig Zugang zu der Information über die Scheidung und man kann nicht von Jones erwarten, dass er sich die Ohren zuhält, wenn er sich in der Nähe von privaten Gesprächen befindet. Aber die Verwendung der mitgehörten Information in Gesprächen mit Freunden verletzt Smith's Privatsphäre. Nur weil Jones Zugang zu den Informationen bekommen hat, bedeutet das nicht, dass er die fraglichen Informationen nutzen oder verbreiten darf.[630]

Acquisti et al. definieren informationelle Privatsphäre folgendermaßen: „(...) privacy as the selective, self-interested opening and

625 Rössler 2004:8
626 Ebd.:8f.
627 Moore bezieht sich auch auf Orte und Körper, aber diese Dimensionen sind hier nicht relevant, da es in dieser Arbeit ausschließlich um informationelle Privatsphäre geht (Moore 2010:27).
628 Moore 2010:28
629 Rachels 1975:333; Moore 2010:30
630 Moore 2010:31

closing of the self to others."[631] Es geht ihrer Meinung nach nicht nur um den Rückzug und das Verstecken der eigenen Informationen, sondern auch um das selektive Öffnen vor anderen aus Eigeninteresse. Acquisti et al. erklären unsere Fähigkeit zu entscheiden, wann wir uns öffnen und wann verschließen, mit Hilfe eines evolutionären Ansatzes. Denn ursprünglich ging es darum, die Anwesenheit anderer Menschen wahrzunehmen und angemessen darauf zu reagieren. Diese Fähigkeit, „sensorial cues" richtig zu interpretieren, kann Menschen vor Risiken schützen und ihnen Vorteile bringen. So kann ein Individuum seine Interessen und Präferenzen mit einem „marketer" teilen, um passende Kaufangebote zu bekommen (Vorteil) und gleichzeitig verheimlichen, wieviel es bereit ist dafür zu zahlen, um Preisdiskriminierung zu vermeiden (Schutz vor Risiken).[632]

Die Privatsphäre einer Person kann laut Mainz und Uhrenfeld nur dann verletzt werden, wenn der Verlust der Kontrolle über ihre Informationen auf die Handlung(en) eines anderen Akteurs zurückzuführen ist. Denn wenn man zum Beispiel sein Tagebuch im Café liegen lässt, wurde einem die Kontrolle über die Informationen darin nicht durch andere verwehrt, solange es nicht gelesen wird.[633] Daraus ergibt sich der folgende Kontroll-Ansatz: „For any agent A to have her right to privacy violated, there is a necessary and sufficient condition that must be satisfied: Agent A has involuntarily lost Negative Control over the access to personal information P about agent A, *due to action(s) of agent B, of which B is responsible*."[634] Die Privatsphäre einer Person ist also nur dann verletzt, wenn sie unfreiwillig die Negative Kontrolle[635] über den Zugang zu persönlichen Informationen über sich verloren hat, und zwar aufgrund von Handlungen eines anderen Akteurs, für die dieser Akteur auch verantwortlich ist.

Nach dem Kontroll-Ansatz ist die Privatsphäre einer Person intakt, wenn sie den Zugang zu Informationen über sich gewähren

631 Acquisti et al. 2022:270
632 Ebd.
633 Mainz, Uhrenfeldt 2021:298
634 Ebd.:298; Die Autoren formulieren auch den Zugangs-Ansatz auf die Weise um, aber verwerfen ihn auf Grund eines Gegenbeispiels.
635 Negative Kontrolle bezeichnet die Möglichkeit andere am Zugang zu den eigenen Informationen zu hindern (Mainz, Uhrenfeldt 2021:293). Was genau damit gemeint ist, wird im folgenden Teilkapitel unter 6.2.2.1 erklärt.

oder verweigern kann, indem sie Kontrolle über den Zugang zu und die Benutzung von den eigenen Informationen hat. Verletzt wird die Privatsphäre der Person hingegen dann, wenn sie die Kontrolle über den Zugang zu ihren Informationen aufgrund von Handlungen eines anderen Akteurs verloren hat. Somit ermöglicht es diese Definition von Privatsphäre nach dem Kontroll-Ansatz zu bestimmen, wann Privatsphäre intakt und wann verletzt ist. Außerdem lässt sich Raum für individuelle Intuitionen, weil der Zugang zu den Informationen selbstständig gewährt oder verweigert werden kann. Um genauer zu verstehen, was mit Kontrolle gemeint ist, werden im Folgenden verschiedene Definitionen von Kontrolle vorgestellt und diskutiert.

6.2.2.1 Definition von Kontrolle

Mainz und Uhrenfeldt unterscheiden zwischen Positiver, Negativer und Republikanischer Kontrolle: Positive Kontrolle über den Zugang zu relevanten Informationen hat man dann, wenn man versucht oder versuchen könnte, jemand anderem tatsächlichen Zugang zu den Informationen zu gewähren und dies gelingt. Negative Kontrolle über den Zugang zu relevanten Informationen hat man hingegen dann, wenn man in der Lage ist, jemand anderen, der versucht, sich Zugang zu verschaffen, am Zugang zu den Informationen zu hindern.[636] Und Republikanische Kontrolle über den Zugang zu relevanten Informationen hat man dann, wenn andere gar nicht erst die Möglichkeit haben, Zugang zu persönlichen Informationen zu erhalten.[637]

Republikanische Kontrolle

Das Problem der Republikanischen Kontrolle ist ihre Strenge. Denn sie führt zu der Annahme, dass die Privatsphäre nur dann intakt ist, wenn andere gar nicht erst die Möglichkeit haben, Zugang zu persönlichen Informationen zu erhalten. Das würde allerdings bedeuten, dass die Privatsphäre auch verletzt wird, wenn man anderen Zugang zu den eigenen Informationen gewährt. Dadurch würde eine Definition der Privatsphäre als Republikanische Kontrolle über

636 Mainz, Uhrenfeldt 2021:293
637 Ebd.

Zugang zu und Benutzung von den eigenen Informationen keinen Raum lassen für individuelle Intuitionen und dementsprechend eins meiner Kriterien nicht erfüllen.

Positive Kontrolle

Laut Mainz und Uhrenfeldt hat man dann Positive Kontrolle über den Zugang zu relevanten Informationen, wenn man versucht (oder versuchen könnte), jemand anderem tatsächlichen Zugang zu den Informationen zu gewähren und dies gelingt.[638] Sie verdeutlichen die Bedeutung der Positiven Kontrolle durch das „Too much info" Beispiel: Smith erzählt seinem Kollegen Jones persönliche Details über sein Sexleben, aber Jones hält sich die Ohren zu und hört die Informationen nicht. Wenn man Kontrolle als Positive Kontrolle interpretiert, hat Jones das Recht von Smith auf Privatsphäre verletzt, indem er seine Ohren zuhält, da Smith dadurch die Positive Kontrolle über den Zugang zu den Informationen verliert.[639] Deshalb verwerfen Mainz und Uhrenfeldt die Positive Kontrolle und auch ich behaupte, dass das Gelingen des Teilens von Informationen keine Bedingung für das Bestehen der Privatsphäre sein kann.

Positive Kontrolle ist trotzdem wichtig, weil man ohne das Gelingen des Teilens von Informationen zum Beispiel keinen Wohngeldantrag stellen könnte. Allerdings hat das dann nichts mehr mit der Privatsphäre der Informationen teilenden Person zu tun, sondern mit erfolgreicher Kommunikation, welche notwendig ist, um bestimmte Zwecke zu erreichen. In Bezug auf die Privatsphäre bedeutet die Möglichkeit Zugang zu gewähren, dass ich Informationen teilen darf, aber das bedeutet nicht, dass ich bestimmen kann, ob die Informationen auch bei dem vorgesehenen Empfänger ankommen.

Negative Kontrolle

Laut Mainz und Uhrenfeldt kann nur der Verlust der Negativen Kontrolle plausibel die Privatsphäre verletzen.[640] Man hat Negative Kontrolle, wenn man andere am Zugang zu seinen Informationen hindern kann – wichtig ist den Autoren dabei, dass auch wirklich jemand anderes versucht sich Zugang zu den Informationen zu

638 Mainz, Uhrenfeldt 2021:293
639 Ebd.:294
640 Ebd.

verschaffen. Das gilt auch für die Verletzung der Privatsphäre: „In order for Negative Control to be lost, someone must *attempt* to get access (...)."[641] Menges argumentiert hingegen, dass die Fähigkeit, andere am Zugang zu hindern, durchaus verringert werden kann, ohne dass jemand überhaupt versucht, Zugang zu erhalten. Wenn man zum Beispiel betäubt ist durch K.O.-Tropfen, ist man nicht in der Lage, andere am Zugang zu hindern und einen zu berühren. Unabhängig davon, ob jemand versucht, tatsächlich Zugang zu bekommen und einen zu berühren, ist man nicht in der Lage dazu, den Zugang zu verhindern: „(...) they lack negative control over whether others touch them even if others do not attempt to touch them."[642] In diesem Fall stimmt es also nicht, dass jemand versuchen muss, tatsächlich Zugang zu bekommen, damit die betroffene Person die negative Kontrolle verliert.

Ich denke jedoch man muss sich hier die Frage stellen, warum die betroffene Person nicht in der Lage ist, andere am Zugang zu hindern, und somit die Negative Kontrolle verloren hat: die K.O.-Tropfen sind der Grund. Deshalb behaupte ich Mainz und Uhrenfeldt würden auf Menges Einwand erwidern, dass es darauf ankommt, ob man sich selber mit K.O.-Tropfen betäubt hat oder ob sie einem von jemand anderem verabreicht wurden. Im letzteren Fall hätte die betäubt Person unfreiwillig die Negative Kontrolle über den Zugang zu den eigenen Informationen verloren, und zwar aufgrund von Handlungen einer anderen Person. Allein das Betäuben der Person könnte dann schon als versuchter Zugang gewertet werden.[643] Denn dem negativen Kontroll-Ansatz von Mainz und Uhrenfeldt zufolge ist die Privatsphäre einer Person nur dann verletzt, wenn sie unfreiwillig die negative Kontrolle über den Zugang zu persönlichen Informationen über sich verloren hat, und zwar aufgrund von Handlungen eines anderen Akteurs, für die dieser Akteur auch verantwortlich ist.[644]

641 Mainz, Uhrenfeldt 2021:300

642 Menges 2022:701f.

643 Das Betäuben einer Person durch die Verabreichung von K.O.-Tropfen verletzt diese Person natürlich auch in anderer Hinsicht, zum Beispiel als Körperverletzung. Aber es ist unter anderem auch die Privatsphäre betroffen.

644 Mainz, Uhrenfeldt 2021:298

Für meine Kriterien heißt das, dass man mit einer Definition von Privatsphäre als Negative Kontrolle über Zugang zu und Benutzung von persönlichen Informationen bestimmen könnte, wann Privatsphäre intakt und wann verletzt ist. Denn die Privatsphäre einer Person ist demnach intakt, solange die Person in der Lage ist, jemand anderen, der versucht, sich Zugang zu verschaffen, am Zugang zu den Informationen zu hindern.[645] Und die Privatsphäre einer Person ist verletzt, sobald die Person einen Zugangsversuch aufgrund von Handlungen anderer Akteure nicht verhindern kann.[646] Diese Definition versperrt im Prinzip nicht den Raum für individuelle Intuitionen, da die Person sich Freunden mitteilen könnte – das scheint keinen Einfluss auf die Privatsphäre zu haben. Dadurch, dass die Privatsphäre als Negative Kontrolle definiert wird, ist allerdings das aktive Teilen von Informationen auch kein Teil der Privatsphäre.

6.2.2.2 Wahl-Kontrolle

Um einen Kompromiss zu finden zwischen dem positiven Gewähren von Zugang und dem negativen Verhindern von Zugang, definiere ich Kontrolle als die Möglichkeit zur Wahl: Das Individuum hat die Kontrolle darüber, mit wem es welche Informationen zu welchem Zweck teilt. Mit der Bezeichnung „Wahl-Kontrolle" orientiere ich mich an dem Control-as-Choice view aus der Literatur, welcher laut Menges impliziert, dass Privatsphäre die Wahl ermöglicht, ob andere Zugang haben oder nicht. Die Vertreter:innen des Kontroll-Ansatzes behaupten laut Menges, dass Privatsphäre im Wesentlichen eine Wahl zwischen Optionen beinhaltet.[647] Dementsprechend stützt sich die Wahl-Kontrolle aus meiner Sicht weder auf rein Positive oder Negative Kontrolle, noch auf Republikanische Kontrolle.[648] Vielmehr vereint sie die Positive und Negative Kontrolle, weil Wahl-Kontrolle bedeutet, dass man in jeder Situation wählen kann, ob man Zugang gewähren oder verhindern möchte.

Ein ähnliches Verständnis von Kontrolle ist das, was Menges „leeway control" (Spielraumkontrolle) nennt: „(...) the ability to

645 Mainz, Uhrenfeldt 2021:293
646 Ebd.:298
647 Menges 2022:693ff.
648 Mainz, Uhrenfeldt 2021:293

do otherwise (...)."[649] Dem Begriff der Spielraumkontrolle zufolge beinhaltet der Schutz der Privatsphäre die Möglichkeit, selbst zu entscheiden, ob andere von bestimmten Informationen erfahren oder Zugang zu ihnen haben sollen oder nicht.[650] Der Begriff der Spielraumkontrolle unterscheidet sich von dem Begriff der Wahl-Kontrolle in zwei Punkten: Erstens geht es bei der Wahl-Kontrolle nicht darum, dass man Kontrolle darüber hat, was andere erfahren – denn hier ist nicht rein Positive Kontrolle gemeint. Vielmehr geht es darum, welche Informationen man bereit ist mit wem zu teilen und zu welchem Zweck. Deshalb ist zweitens die Benutzung der Informationen aus Moores Definition von Privatsphäre wichtig für den Begriff der Wahl-Kontrolle: Nur weil man Zugang zu gewissen Informationen gewährt bekommen hat, heißt das nicht, dass man sie zu jedem Zweck benutzen kann.[651]

Bei der Definition von Privatsphäre orientiere ich mich an dem Kontroll-Ansatz von Moore: „A right to privacy is a right to control access to, and uses of (...) personal information."[652] Da in dieser Arbeit davon ausgegangen wird, dass ein solches moralisches Recht auf Privatsphäre nur aus einer Pflicht zum Schutz der Privatsphäre resultieren könnte[653] und diese Pflicht erst unter 6.3 begründet wird, möchte ich den Begriff des Rechts umgehen – auch um klarer über den Begriff der Privatsphäre sprechen zu können. Nun lässt sich aber das „Recht" nicht einfach aus der Definition von Moore rausstreichen. Aus meiner Sicht lässt sich die Formulierung aber übersetzen in: Wer Privatsphäre hat, hat Kontrolle über Zugang zu und Benutzung von persönlichen Informationen. Moore selbst scheint Kontrolle als negative Kontrolle zu interpretieren, denn seiner Meinung nach ist der Zustand der Privatsphäre gegeben, wenn eine Person sich freiwillig von ihren Mitmenschen abgrenzt und den Zugang zu ihren persönlichen Informationen einschränkt.[654] Ich möchte jedoch dieses negative Verständnis von Kontrolle ersetzen durch das Verständnis von Kontrolle als Wahl darüber, mit wem man welche Informationen zu welchem Zweck teilt.

649 Menges 2020a:34
650 Ebd.:45
651 Moore 2010:28
652 Ebd.:27
653 Mohr 2010:77
654 Mainz, Uhrenfeldt 2021:294; Moore 2010:26

Intakte Privatsphäre

Nach dieser Definition ist die Privatsphäre einer Person intakt, wenn sie wählen kann, wem sie Zugang zu welchen Informationen gewährt oder verweigert, und wenn sie zudem wählen kann, zu welchem Zweck ihre Informationen benutzt werden. Bei dieser Wahl kann der Kontext des Informationsflusses eine wichtige Rolle spielen. Beispielsweise berichtet ein Individuum seiner Ärztin Details über seinen Gesundheitszustand, die es selbst mit guten Freunden nicht teilen würde. Dafür erzählt das Individuum guten Freunden von einem Streit, von dem es wiederum seine Ärztin nicht in Kenntnis setzen würde. Was die Person mit wem teilt und zu welchem Zweck ist aber nicht wie bei der Technologiephilosophin Helen Nissenbaum[655] an Informationsnormen gebunden, sondern erfolgt nach individueller Intuition – im besten Fall gekoppelt mit Wissen über die Konsequenzen.

Nicht immer sind sich die betroffenen Individuen allen Konsequenzen bewusst. So kann zum Beispiel beim Gewähren des Zugangs zu Informationen die Wahl-Kontrolle vermindert sein durch zum Beispiel Drogen, psychische Probleme oder Unwissenheit. Wenn diese verminderte Kontrolle ausgeübt wird, wird dadurch auch die Privatsphäre vermindert.[656] Aber bei unverminderter Kontrolle bedeutet das Preisgeben von persönlichen Informationen nicht, dass dadurch die Privatsphäre aufgegeben wurde. So schreibt Menges: „(...) we can share private information with a good friend without giving up privacy."[657] Wenn wir uns bewusst anderen Menschen mitteilen, beziehen wir die Anderen in unseren Bereich der Privatsphäre ein, ohne unsere Privatsphäre zu beeinträchtigen.[658]

Es gibt Kritiker, die einwenden, dass die Kontrolle allein nicht die Privatsphäre ausmacht – es müsse auch um einen Grad der Abgeschiedenheit gehen.[659] Wenn eine Person, wie im *Voluntary Divulgence Case* unter 6.1, Fremden von ihren Problemen erzählt, hat sie Zugang zu ihren Informationen gewährt und man kann nicht davon sprechen, dass sie dadurch ihre Abgeschiedenheit för-

655 Nissenbaum 2010:188
656 Menges 2022:702
657 Ebd.:702f.
658 Inness 1992:46
659 Masur et al. 2018:5f.

dert. Angenommen, sie hat sich bewusst dazu entschieden, diese Informationen zu teilen, so hat sie ihre Wahl-Kontrolle über den Zugang zu diesen Informationen ausgeübt. Sie hat sich nicht für den Rückzug entschieden, sondern kontrolliert ihre Informationen geteilt. Das heißt aber nicht, dass sich die Person auch zukünftig dafür entscheiden wird, ihre Probleme mit Fremden zu teilen. Die Entscheidung, Informationen kontrolliert zu teilen, hebt nicht die Möglichkeit zum Rückzug auf. Die Person hat durch das einmalige Teilen ihrer privaten Informationen ihre Privatsphäre nicht verletzt – sie wurde nur temporär ausgeweitet. Nach der Ansicht des Psychologen Irwin Altman schwankt das gewünschte Maß an Privatsphäre je nach Situation und Interaktion, deshalb passen die Menschen ihre eigenen Grenzen immer wieder an.[660]

Hier kommt jedoch eine zeitliche Dimension hinzu: Während eine Person freiwillig und nüchtern fremden Menschen von ihren persönlichen Problemen und dem regelmäßigen Alkoholkonsum erzählt, hat sie die Wahl-Kontrolle über diese Informationen. Aber sobald die fremden Menschen über die persönlichen Informationen verfügen, können die Fremden die Informationen weiterverbreiten und die erzählende Person verliert die Kontrolle darüber, was mit ihren Informationen in Zukunft passiert. Der Gedanke findet sich auch bei Menges: „(...) exercising full control by telling strangers about personal information can and often does threaten our future privacy with regard to the information."[661] Wenn die erzählende Person den Fremden Zugang zu Informationen über ihre persönlichen Probleme gewährt, stimmt sie dadurch nicht automatisch zu, dass die Fremden ihre Informationen benutzen dürfen, indem diese sie zum Beispiel weiterverbreiten. Nach dem Gewähren des Zugangs zu Informationen ist es jedoch möglich, dass die Informationen benutzt werden – dessen sollte man sich beim Teilen der Informationen bewusst sein.

Verletzung der Privatsphäre

Die Privatsphäre einer Person wird verletzt, wenn sie nicht mehr wählen kann, wem sie Zugang zu welchen Informationen gewährt oder verweigert, oder wenn sie nicht mehr wählen kann, zu wel-

660 Altman 1976:13f.; Dienlin 2014:107
661 Menges 2022:703

chem Zweck ihre Informationen benutzt werden. Ich stimme Mainz und Uhrenfeld zu und argumentiere, dass für diese eingeschränkte oder fehlende Wahl-Kontrolle immer Außenstehende verantwortlich sein müssen, da man seine Privatsphäre nicht selbst verletzen kann.[662] Eine Person kann Drogen konsumieren und dadurch ihre Möglichkeit zur Wahl-Kontrolle selber einschränken – denn man kann womöglich im berauschten Zustand nicht mehr so wählen, wie man das im nüchternen Zustand tun würde und könnte sich im Nachhinein für das bereitwillige Gewähren von Zugang zu den eigenen Informationen schämen. Solange sich die Person aber freiwillig für das Konsumieren der Drogen entschieden hat und sich der Konsequenzen bewusst war, hat sie im Vorhinein ihre Kontrolle über ihren eigenen Bewusstseinszustand ausgeübt.

Ob die Privatsphäre des Individuums verletzt wird, hängt dementsprechend vom Verhalten der Außenstehenden ab. Außenstehende können die Privatsphäre eines Individuums verletzen, indem sie entweder *bewusst* versuchen auf persönliche Informationen des Individuums zuzugreifen oder diese zu benutzen, ohne dass das Individuum darüber die Wahl-Kontrolle hat, oder Außenstehende greifen tatsächlich, womöglich auch *aus Versehen*, auf persönliche Informationen des Individuums zu oder benutzen diese, ohne dass das Individuum darüber die Wahl-Kontrolle hat.

Ein tatsächlicher, aber unbeabsichtigter Zugriff ohne die Kontrolle des Individuums lässt sich durch folgendes Beispiel beschreiben: Ein Kind benutzt eine öffentliche Toilette und schließt die Tür nicht ab, weil es Sorge hat, das Schloss alleine nicht wieder öffnen zu können. Eine weitere Toilettennutzerin öffnet die Tür, in der Annahme, dass die Toilette nicht besetzt ist. Die Toilettennutzerin sieht das Kind, entschuldigt sich und schließt die Tür schnell wieder. Sie verletzt in dem Fall die Privatsphäre des Kindes, weil sie aus Versehen auf persönliche Informationen des Kindes zugreift. Das Kind ist dieses Risiko eingegangen, als es die Tür nicht abgeschlossen hat. Es hat also bewusst etwas von seiner Wahl-Kontrolle über den Zugang zu und die Benutzung von persönlichen Informationen aufgegeben, weil es das Abschließen vermeiden wollte. So kann es manchmal bei einer Abwägung von möglichen Risiken dazu kommen, dass der Schutz der Privatsphäre nicht die oberste Priorität hat. Das Indivi-

662 Mainz, Uhrenfeldt 2021:298

duum kann sich jederzeit dafür entscheiden, etwas Kontrolle über den Zugang zu und die Benutzung von persönlichen Informationen abzugeben und zu riskieren, dass Außenstehende dann tatsächlich auf persönliche Informationen des Individuums zugreifen. So ein tatsächlicher, aber unbeabsichtigter Zugriff auf persönliche Informationen sollte im Datenökosystem nicht möglich sein, weil die Daten so gespeichert sein sollten, dass niemand zufällig darauf stoßen kann.

Eine tatsächliche Benutzung persönlicher Informationen ohne die Kontrolle des Individuums passiert zum Beispiel dann, wenn Menschen, denen man sich anvertraut, das Anvertraute weitergeben. Das Weitergeben der Informationen verletzt die Privatsphäre, weil die ursprüngliche Quelle der Informationen dann keine Wahl-Kontrolle über Zugang zu und Benutzung von ihren persönlichen Informationen mehr hat. Das heißt, solange man nicht gebeten wurde, persönliche Informationen von anderen weiterzugeben, stellt das eine Verletzung der Privatsphäre dieser anderen Personen dar. Hier lässt sich eine klare Parallele zu der Zweckentfremdung von personenbezogenen Daten erkennen. Die Wahl-Kontrolle des betroffenen Individuums ist nur intakt, wenn die Daten ausschließlich für vereinbarte Zwecke verwendet werden.

Der Versuch des Zugriffs auf oder der Benutzung von persönlichen Informationen lässt sich durch das Abhörbeispiel illustrieren: Wenn jemand das Telefon eines anderen abhört, aber keine Informationen bekommt, beeinträchtigt er trotzdem allein durch seinen Versuch die Privatsphäre des anderen.[663] Übertragen auf das Datenökosystem bedeutet dies, dass allein der Versuch, auf Daten zuzugreifen, ohne dass das betroffene Individuum darüber die Kontrolle hat, die Privatsphäre des Individuums verletzt.

Fahrlässigkeit und Beihilfe

Wie unter 6.2.1 dargestellt, verletzen laut Lundgren Außenstehende die Privatsphäre eines Individuums außerdem dann, wenn sie das Individuum einem substanziell größeren Risiko aussetzen, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift.[664] Wenn zum Beispiel ein Arzt während der Behandlung die

663 Mainz, Uhrenfeldt 2021:299
664 Lundgren 2021b:387

Tür des Behandlungszimmers offenstehen lässt, ist das Risiko substanziell größer, dass eine andere außenstehende Person vorbeiläuft, hineinschaut und (womöglich auch aus Versehen) auf persönliche Informationen der Patient:innen zugreift. Der Arzt könnte die Tür des Behandlungszimmers absichtlich offenstehen lassen, um das Risiko zu erhöhen, dass jemand Unbefugtes hineinschaut. Allerdings würde ich ein solches Fehlverhalten keinem Arzt unterstellen wollen. Trotzdem kann die Situation entstehen durch fahrlässiges Handeln, weil die Tür zum Beispiel aus Eile nicht vollständig geschlossen wurde. Eine solche fahrlässige Handlung bezeichne ich als Beihilfe zur Privatsphärenverletzung und stimme Lundgren damit zu, dass der Anspruch auf Privatsphäre auch einen Schutz vor bestimmten Risiken für die Privatsphäre umfasst.[665] Das heißt im Kontext von Datenökosystemen, dass Datengenerierende aus ethischer Sicht nicht fahrlässig mit den Daten der Datenquellen umgehen sollten.

Lundgren hat diese Spezifikation der Privatsphärenverletzung zwar aus Sicht des Zugangs-Ansatzes aufgestellt, aber ich argumentiere, dass sie sich auf die Sicht des Kontroll-Ansatzes übertragen lässt. Denn auch aus Sicht des Kontroll-Ansatzes kann der tatsächliche Zugang zu persönlichen Informationen problematisch sein – das läge dann allerdings nicht an dem Zugang selbst, sondern an der fehlenden Kontrolle über den Zugang. Ein Individuum einem substanziell größeren Risiko aussetzen, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift bedeutet also für den Kontroll-Ansatz nicht nur, dass der Zugang für Außenstehende erleichtert wird, sondern auch, dass die Kontrolle des Individuums beschränkt wird oder die schon beschränkte Kontrolle nicht berücksichtigt wird.

Laut Lundgren verletzen Außenstehende die Privatsphäre eines Individuums außerdem dann, wenn sie dafür sorgen, dass das Individuum glaubt, dass oder unsicher ist, ob seine privaten Angelegenheiten dem Risiko des Zugriffs durch eine andere außenstehende Person ausgesetzt sind.[666] Wenn Außenstehende einen solchen Glauben oder Unsicherheit bei den betroffenen Individuen auslösen kann das zwar normativ verurteilt werden, aber es stellt allein noch keine Verletzung der Privatsphäre dar. Ich denke, eine intakte Privatsphäre

665 Lundgren 2021b:387
666 Ebd.:389

umfasst nicht das Wissen darüber, ob die persönlichen Informationen dem Risiko des Zugriffs durch eine außenstehende Person ausgesetzt sind. Denn die Verletzung der Privatsphäre geschieht nicht, wenn das Individuum das Gefühl hat, die Kontrolle zu verlieren, sondern wenn die Außenstehenden tatsächlich zugreifen oder versuchen zuzugreifen.

Zusammenfassung

In der Literatur wird nicht klar unterschieden zwischen der Perspektive des Individuums und der Perspektive von Außenstehenden – ich denke das ist ein Fehler. Denn hierdurch wird deutlich, dass nur Außenstehende die Privatsphäre des Individuums verletzen können – dies geschieht durch eine Handlung, die nicht zwingend eine schlechte Intention haben muss. Dank Tabelle 2 wird deutlich, wann die Privatsphäre von Individuen intakt und wann sie verletzt ist, und was Außenstehende vermeiden sollten, um die Privatsphäre der Individuen nicht zu verletzen.

Tabelle 2 Teile der Privatsphäre als Wahl-Kontrolle[667]

Wahl-Kontroll-Ansatz	**Individuum**	**Außenstehende**
Privatsphäre intakt	Das Individuum hat Wahl-Kontrolle über den Zugang zu und die Benutzung von persönlichen Informationen.	Außenstehende versuchen nicht auf persönliche Informationen des Individuums zuzugreifen, ohne dass das Individuum darüber die Wahl-Kontrolle hat, und sie greifen auch nicht tatsächlich, nicht einmal aus Versehen, darauf zu.
Privatsphäre verletzt	Das Individuum hat eine eingeschränkte oder fehlende Wahl-Kontrolle über den Zugang zu und die Benutzung von persönlichen Informationen.	Außenstehende versuchen *bewusst* auf persönliche Informationen des Individuums zuzugreifen, ohne dass das Individuum darüber die Wahl-Kontrolle hat. ODER Außenstehende greifen tatsächlich, womöglich auch *aus Versehen*, auf persönliche Informationen des Individuums zu und benutzen diese, ohne dass das Individuum darüber die Wahl-Kontrolle hat.

667 Eigene Darstellung

Abschließend lässt sich sagen, dass es die Definition der Privatsphäre als Wahl-Kontrolle_ermöglicht zu bestimmen, wann Privatsphäre intakt und wann sie verletzt ist.

Privatsphäre ist intakt, wenn das Individuum die Wahl-Kontrolle hat über Zugang zu und Benutzung von seinen persönlichen[668] Informationen. Das heißt, das Individuum hat die Kontrolle darüber, mit wem es welche Informationen zu welchem Zweck teilt.

Privatsphäre wird verletzt, wenn das Individuum die Kontrolle über Zugang zu und die Benutzung von persönlichen Informationen verliert. Außenstehende können diesen Kontrollverlust verursachen, indem sie bewusst versuchen auf persönliche Informationen des Individuums zuzugreifen und diese zu benutzen oder tatsächlich zugreifen, womöglich auch aus Versehen, ohne dass das Individuum darüber die Kontrolle hat. Außerdem können Außenstehende durch fahrlässiges Handeln Beihilfe zur Privatsphärenverletzung leisten, indem sie das Individuum einem substanziell größeren Risiko aussetzen, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift.

Außerdem lässt die Definition von Privatsphäre als Wahl-Kontrolle Raum für individuelle Intuitionen, da die Individuen nach eigenem Ermessen Zugang gewähren und verweigern können, und erfüllt damit all meine Kriterien. Ob alle Individuen ein Recht auf Privatsphäre im Sinne ihrer Wahl-Kontrolle haben, soll im folgenden Kapitel erörtert werden.

6.3 Moralische Pflicht zum Schutz der Privatsphäre und resultierendes Recht auf Privatsphäre

Eine Pflicht zum Schutz der Privatsphäre sowie ein daraus abgeleitetes Recht auf Privatsphäre würden vor Handlungen schützen, die das Risiko einer erheblichen Beeinträchtigung der Privatsphäre mit sich bringen.[669] Vor welchen Handlungen eine solche moralische Pflicht und das daraus resultierende Recht auf Privatsphäre genau

668 Das *Persönliche* bezieht sich hier darauf, dass die Informationen das Individuum betreffen und es deshalb entscheiden kann, mit wem es welche Informationen zu welchem Zweck teilt.

669 Lundgren 2021a:672

schützt, hängt je nach Autor:innen von der jeweiligen Definition der Privatsphäre ab. In dieser Arbeit soll die moralische Pflicht zum Schutz der Privatsphäre die Datenquellen davor bewahren die Wahl-Kontrolle über ihre Daten zu verlieren. Das Ziel dieses Kapitels ist es, eine moralische Pflicht zum Schutz der Privatsphäre der Datenquellen zu begründen und zu verstehen, wie daraus ein Recht auf Privatsphäre resultieren kann.

Eine solche moralische Pflicht zum Schutz der Privatsphäre lässt sich aus Kants erster Formulierung des Kategorischen Imperativs ableiten.[670] Denn wenn eine Handlungsmaxime weder gedacht noch gewollt werden kann, ist es eine vollkommene Pflicht, diese Handlung zu vermeiden. Eine Handlungsmaxime, die zwar denkbar ist, aber nicht widerspruchsfrei gewollt werden kann, etabliert eine unvollkommene Pflicht.[671] Es ist zwar denkbar, dass die Privatsphäre von Datenquellen verletzt wird, um andere Ziele zu verwirklichen, aber es kann nicht gewollt werden – das etabliert eine unvollkommene Pflicht zum Unterlassen der Verletzung der Privatsphäre. Diese Pflicht sowie eine damit einhergehende Pflicht zum Schutz der Privatsphäre lassen sich begründen, indem zum einen erklärt wird, warum die Verletzung der Privatsphäre von Individuen vermieden werden sollte, und zum anderen aufgezeigt wird, warum eine intakte Privatsphäre erstrebenswert ist.

Die Vermeidung von Privatsphärenverletzungen ist ein guter Grund, warum die Privatsphäre schützenswert ist. Der Gedanke findet sich auch bei O'Neill: „Securing privacy matters because it can protect individuals from interference (…)".[672] Der Schutz vor solchen Eingriffen ist deshalb notwendig, weil Verletzungen der Privatsphäre die Möglichkeit zur Erpressung schaffen können.[673] Denn wenn Unbefugte an sensible Gesundheitsdaten gelangen, könnten sie zum Beispiel Menschen mit einer psychischen Diagnose erpressen. Der Schutz der Privatsphäre erlaubt es Personen, ihre psychische Diagnose nur mit denjenigen Personen zu teilen, denen sie vertrauen.

670 Kant 2008:53 [421]
671 Ebd.:55–58 [422–424]
672 O'Neill 2021:79
673 Ebd.:79

Laut O'Neill ist es möglich, dass durch die Verbreitung von persönlichen Informationen Schaden angerichtet wird: „(…) in connected societies where information trading is both efficient and nearly without cost, disclosure of personal information opens individuals up to certain risks—such as being controlled by entities with their own agendas."[674] Wenn Informationen über eine Person für die Öffentlichkeit zugänglich gemacht werden, dann wird diese Person schlechter gestellt im Sinn eines erhöhten Risikos für zum Beispiel staatliche Kontrolle oder kommerzielle Ausbeutung.[675]

Persönliche Informationen können auch von Regierungen zur Unterdrückung der Staatsbürger:innen und zur Erweiterung der eigenen Macht benutzt werden.[676] Zum Beispiel gibt es in China über jeden Menschen eine Akte namens „Dangan", die ab der Grundschule geführt wird und alle möglichen persönlichen Informationen enthält. Insbesondere Einträge zur politischen Einstellung können die beruflichen Möglichkeiten und die Reisefreiheit der Menschen beeinflussen.[677] Gleichermaßen können Wirtschaftsunternehmen persönliche Informationen dazu benutzen, Kunden mit einer Flut an Werbung zu überfordern oder auch ihre Angestellten zu kontrollieren.[678] Insbesondere, wenn diese Informationen digital gespeichert werden, erhöht sich das Risiko, weil Daten im digitalen Raum dauerhaft zugreifbar bleiben.[679] Deshalb werden den Menschen laut Moore durch die Digitalisierung, die Gefahr der Manipulation und die Möglichkeit der Verbreitung von persönlichen Informationen als Risiken auferlegt, welche ungerechtfertigt Schaden anrichten können.[680]

Der Schutz der Privatsphäre verhindert also, dass Menschen diskriminiert werden beispielsweise wegen einer bestimmten politischen Einstellung, und dadurch wiederum erlaubt sie es den Menschen, ihre Rechte wahrzunehmen, zum Beispiel zu demonstrieren. Wenn eine Regierung wüsste, welche politische Einstellung die Staatsbürger:innen haben und wo sie sich aufhalten, dann könnten

674 O'Neill 2021:88
675 Ebd.:86, 91
676 Moore 2010:88
677 Moore 2010:89
678 Ebd.:89f.
679 Ebd.:90f.
680 Ebd.:98

sie das Versammeln von Gruppen zu Demonstrationen verhindern. Die Privatsphäre schützt vor solchen Eingriffen und untermauert damit grundlegende Menschenrechte.[681] Außerdem verhindert der Schutz der Privatsphäre, dass sich die betreffenden Personen für die offenbarten Informationen schämen, sich verfolgt fühlen oder andere mentale Beeinträchtigungen erleiden.[682]

Zudem ermöglicht eine intakte Privatsphäre auch mehr Autonomie, da die Menschen in privaten Räumen nicht von anderen bewertet werden und dementsprechend unabhängiger entscheiden können.[683] Rössler behauptet, dass „(...) Autonomie nicht, oder jedenfalls nicht in all ihren und nicht in ihren relevantesten Aspekten, lebbar ist ohne den Schutz von Privatheit, ohne die Unterscheidung zwischen privaten und öffentlichen Dimensionen oder Bereichen des Lebens."[684] Einerseits beschützt die Privatsphäre also das, was der Philosoph Isaiah Berlin negative Freiheit nennt – die Abwesenheit von Eingriffen.[685] Andererseits unterstützt die Privatsphäre aber auch die von O'Neill sogenannten „capacities to act"[686], ermöglicht dadurch also die positive Freiheit.[687] Der Schutz der Privatsphäre schließt also auch andere Werte ein.[688]

Laut Moore ist Privatsphäre für das menschliche Wohlbefinden unerlässlich, da die Fähigkeit, sich zurückzuziehen, für Menschen eine Lebensnotwendigkeit und gleichzeitig die Gesellschaft anderer Menschen ein menschliches Grundbedürfnis ist.[689] Um beidem gerecht zu werden, öffnen sie ihre Privatsphäre selektiv: „Somewhere on the continuum between one's innermost thoughts that are never shared and what one offers up for anyone's consideration we switch from the private to the public."[690] Daraus schlussfolgert Moore, dass die Fähigkeit, den Zugang zu und die Nutzung von persönlichen Informationen zu regeln, für das menschliche Wohlbefin-

681 Bernal 2016:252
682 Price, Cohen 2019:41
683 Masur et al. 2018:6; Westin 1967, S. 35–43
684 Rössler 2003:30
685 Berlin 1997:195
686 O'Neill 2021:79
687 Berlin 1997:196, 202f.
688 Barocas, Nissenbaum 2014:49
689 Ebd.:48f.; 86f.
690 Moore 2010:87

den unerlässlich ist.[691] In ähnlicher Weise argumentiert Fried, dass der Entzug der Privatsphäre die Möglichkeit, Intimität zu gewähren, zerstört und dadurch die wesentliche Dimension von Liebe und Freundschaft unmöglich macht, sowie die Fähigkeit des Subjekts, Vertrauensbeziehungen einzugehen, untergräbt. Da aber intime Beziehungen der Liebe und der Freundschaft dem menschlichen Leben einen großen Teil des positiven Wertes verleihen und die Privatsphäre dafür den Rahmen bildet, ist ihr Schutz notwendig.[692]

Moore betont weiter, dass Privatsphäre nicht nur schützenswert, sondern auch wertvoll sei, da sie für das Überleben und die geistige Gesundheit notwendig ist.[693] Daraus ergibt sich ein Anspruch auf Kontrolle über persönliche Informationen – zumindest solange dadurch niemand geschädigt wird. Da Informationen von mehreren Personen gleichzeitig besessen, konsumiert oder betrachtet werden können, ohne einander zu schaden[694], entstehen moralische Nutzungsansprüche an die eigenen Informationen, sofern niemand dadurch schlechter gestellt wird.[695]

Die Verletzung der Privatsphäre stellt also insbesondere deshalb ein unerwünschtes Ereignis dar, weil sie mit einer Risikoauferlegung gleichzusetzen ist – es entsteht das Risiko ungewollter Eingriffe und Diskriminierung sowie das Risiko fehlender Autonomie und Intimität. Ich stimme also Moore zu, dass den betroffenen Individuen durch die Verletzung der Privatsphäre ein unzulässiges Risiko auferlegt wird.[696] Für Moore ist zwar die Schlussfolgerung aus dieser Argumentation, dass es ein Recht auf Privatsphäre geben sollte, aber die Argumente für eine intakte Privatsphäre und gegen die Verletzung der Privatsphäre können ebenfalls zu der Schlussfolgerung führen, dass es eine Pflicht zum Schutz der Privatsphäre geben sollte. Da in dieser Arbeit davon ausgegangen wird, dass ein Recht auf Privatsphäre nur aus einer Pflicht zum Schutz der Privatsphäre resultieren kann[697], ist hier die Schlussfolgerung, dass es eine Pflicht zum Schutz der Privatsphäre geben sollte. Die moralische Pflicht

691 Moore 2010:21; 56
692 Fried 1968:216
693 Moore 2010:43f., 86f.
694 Ebd.:82, 98
695 Ebd.:84f.
696 Ebd.:97
697 Mohr 2010:77

der Datengenerierenden zum Schutz der individuellen Privatsphäre begründet sich dabei sowohl durch die negativen Konsequenzen, die eine Verletzung der Privatsphäre verursachen kann – es entstehen Risiken, die für das Wohlbefinden der betroffenen Person moralisch relevant sind[698] – als auch durch die positiven Konsequenzen, für die eine intakte Privatsphäre sorgen kann.[699]

Dabei sind aus deontologischer Sicht genau die Handlungsfolgen relevant, welche die moralischen Grundsätze verbieten oder herbeiführen wollen. Somit ist laut Hare aus deontologischer Sicht das Verletzen der Privatsphäre von Datenquellen unter anderem deshalb falsch, weil es Erpressung, staatliche Kontrolle oder kommerzielle Ausbeutung zur Folge haben kann. Der Grundsatz des Kategorischen Imperativs verbietet die Handlungsfolgen der Erpressung, staatlicher Kontrolle oder kommerzieller Ausbeutung, weil die zugrundeliegende Handlungsmaxime nicht gewollt werden kann und somit eine unvollkommene Pflicht entsteht, diese Handlung zu vermeiden. Entsprechend dieses Grundsatzes ist es moralisch relevant, wenn die Folge einer Handlung entweder Erpressung, staatliche Kontrolle oder kommerzielle Ausbeutung ist.[700] Im Prinzip ist es denkbar, dass sich die Menschen gegenseitig erpressen, kontrollieren und ausbeuten. Aber niemand kann es wollen erpresst, staatlich kontrolliert oder kommerziell ausgebeutet zu werden.

Außerdem ist laut Hare das Verletzen der Privatsphäre von Datenquellen deshalb falsch, weil sich Datenquellen ohne intakte Privatsphäre nicht autonom entfalten können und zum Beispiel keine intimen Beziehungen der Liebe und der Freundschaft eingehen können – das schadet ihrem Wohlbefinden und dadurch können sie ihrem Selbstzweck nicht gerecht werden.[701] In dem die Privatsphäre der Datenquellen verletzt wird, also ihre Daten ohne ihre Kontrolle und ihre Zustimmung verwendet werden, werden die Datenquellen als bloßes Mittel gebraucht. Die dritte Formel des Kategorischen Imperativs verbietet aber, dass Datennutzende in dieser Weise als bloßes Mittel gebraucht werden. Diese Konsequenz, Datenquellen die Möglichkeit zu verwehren sich autonom zu entfalten für ihr Wohl-

698 Moore 2010:97 Rössler 2003:3
699 Masur et al. 2018:6; Vgl. Westin 1967, S. 35–43; Fried 1968:216
700 Hare 1993:15
701 Rössler 2003:30; Moore 2010: 21; 56, 87; Fried 1968:216

befinden (und sie dadurch bloß als Mittel zu gebrauchen), ist das, was das Verletzen der Privatsphäre falsch mache.[702] So lässt sich aus den negativen Konsequenzen, die eine Verletzung der Privatsphäre verursachen kann mit der deontologischen Perspektive von Hare eine vollkommene Pflicht zum Schutz der Privatsphäre begründen – denn es ist nicht denkbar, dass die Menschen den Kategorischen Imperativ ignorieren und sich als bloße Mittel gebrauchen.

Aus dieser vollkommenen moralischen Pflicht zum Schutz der Privatsphäre auf Seiten der Datengenerierenden kann ein Primärrecht auf Privatsphäre für Datenquellen resultieren.[703] Der praktische Philosoph Ingmar Persson und der Medizin- und Technnikkethiker Julian Savulescu argumentieren jedoch, es sei unmöglich, dass es ein Recht auf Privatsphäre geben kann. Allerdings legen sie ihrem Argument eine definitorische Voraussetzung zu Grunde, die ich hier widerlegen möchte. Die beiden Autoren definieren ein Recht auf Privatsphäre folgendermaßen: „A right to privacy is conceived as a right that others do not acquire information about us that we reserve for ourselves and selected others."[704] Dabei machen sie – ohne dies begrifflich zu kennzeichnen – eine Unterscheidung zwischen dem passiven Erlangen (acquire) der Informationen und den Mitteln, die zum aktiven Beschaffen (acquire) der Informationen verwendet werden. Ihre These ist, dass wir das Recht gegenüber anderen haben können, dass sie sich von bestimmten Mitteln der Informationsbeschaffung fernhalten, aber nicht von der Informationsbeschaffung an sich.

Aus Perssons und Savulescus Sicht ist das Erlangen von Informationen keine Handlung, sondern ein innerer Zustand, der herbeigeführt wird.[705] Im Verlaufe des Artikels wird klar, dass mit dem Erlangen (acquire) von Informationen das passive Wahrnehmen von Informationen gemeint ist – nicht das aktive Beschaffen. Denn Persson und Savulescu vergleichen das Erlangen von Informationen mit dem Empfinden von Emotionen: „(…) the state of acquiring information is in itself evaluatively neutral."[706] Es geht ihnen also darum, dass es kein Recht zum Schutz gegen das zufällige und unabsichtliche Wahrnehmen von privaten Informationen geben kann.

---

702 Hare 1993:15
703 Mohr 2010:77; Mieth, Bambauer 2016:3
704 Persson, Savulescu 2022:1
705 Ebd.
706 Ebd.:3

Diese These beruht aber auf einem Missverständnis: Das Recht auf Privatsphäre soll kein Recht sein auf den Schutz vor Sinneswahrnehmungen von Mitmenschen, sondern vor deren absichtlicher Informationsbeschaffung. Denn die informationelle Privatsphäre wird nicht mehr vorrangig im direkten Kontakt durch Sinneseindrücke verletzt, sondern online. Im digitalen Raum lassen sich die Mittel nicht mehr so leicht von der Informationsbeschaffung trennen, da man kaum unabsichtlich private Informationen wahrnehmen kann, die nicht geteilt werden sollten. Das Gegenüber könnte im Zoom Meeting aus Versehen die Kamera angeschaltet haben, sodass eine private Situation wahrzunehmen ist. Aber ansonsten hängt die Beschaffung von Informationen meist mit einem bewusst ausgewählten Mittel zusammen.

An dieser Stelle lohnt es sich, die Argumentationen für ein (moralisches) Recht auf Privatsphäre zu betrachten, da hieraus weitere Gründe für eine Pflicht zum Schutz der Privatsphäre ersichtlich werden. Laut Moore gibt es grob zwei Perspektiven auf das Recht zur Privatsphäre: Es gibt in der philosophischen Debatte die Reduktionist:innen, die argumentieren, dass die Privatsphäre von anderen Rechten, wie dem Recht auf Leben, Freiheit und Eigentum, abgeleitet ist und es gibt die der Nicht-Reduktionist:innen, welche die Privatsphäre als mit anderen Rechten oder moralischen Konzepten verwandt betrachten, diese aber als eine eigene Art von Rechten anerkennen.[707]

Reduktionismus

Die bereits erwähnte Moralphilosophin und Metaphysikerin Judith Jarvis Thomson ist der Meinung, dass das Recht auf Privatsphäre ein abgeleitetes Recht ist. Das werde dadurch deutlich, dass man jede Verletzung der Privatsphäre erklären könne, ohne das Recht auf Privatsphäre zu erwähnen. Wenn zum Beispiel jemand beobachtet oder belauscht wird, erscheint das wie eine Verletzung der Privatsphäre. Thomson meint dazu, Menschen hätten das Recht, nicht beobachtet und belauscht zu werden, aber nicht, weil sie ein Recht auf Privatsphäre haben, sondern weil sie Rechte an ihrer Person haben. Wenn jemand gefoltert wird, um ihm persönliche Informationen zu entlo-

707 Moore 2010:14ff.

cken, wird seine Privatsphäre verletzt. Aber laut Thomson muss das nicht mit Hilfe seines Rechtes auf Privatsphäre erklärt werden, sondern mit seinem Recht, nicht verletzt oder geschädigt zu werden. Und wenn sich jemand den Inhalt eines fremden Safes anschaut, ist das laut Thomson falsch, weil jeder das Recht hat, dass sein Eigentum nicht angesehen wird – nicht, weil man ein Recht auf Privatsphäre hat.[708] Laut Thomson lässt sich die Verletzung der Privatsphäre vollständig mit Verstößen gegen andere Individualrechte erklären.[709]

Auch Macnish meint, dass durch den Verlust der Kontrolle über die eigenen Informationen zwar nicht die Privatsphäre verletzt wird, aber ein Risiko für anderweitigen Schaden entsteht. Sobald die Fremde im Café die Kontrolle über die Informationen im Tagebuch übernimmt, entscheidet jemand anderes als die Person selbst, was mit dem Tagebuch und den darin enthaltenen Informationen geschieht. Die Fremde aus dem Café, die über die Informationen im Tagebuch verfügt, kann nun sowohl auf die Informationen zugreifen (d. h. die Privatsphäre der Person verletzen) als auch, unter sonst gleichen Bedingungen, die Informationen wahrscheinlich zu ihrem Vorteil nutzen. Es gibt Anreize für die Fremde aus dem Café, so zu handeln, dass sie davon profitiert und die Tagebuchbesitzerin die Kosten dafür trägt. Es sei rational, unter der Annahme zu handeln, dass sie dies bereits getan hat. Auch wenn die Fremde aus dem Café – die neue Kontrolleurin der Informationen – nicht tatsächlich auf die Informationen zugegriffen hat, sind die Folgen so, als hätte sie es getan. Es handelt sich also um eine Verletzung der Rechte und Interessen, die wir durch Geheimhaltung bestimmter Informationen schützen wollen[710]: „(...) a loss of control may involve harms to the rights that privacy protects."[711] Die Verletzung dieser Rechte und Interessen könne sich wie eine Verletzung der eigenen Privatsphäre anfühlen, auch wenn sie es nicht sei. Denn durch den Verlust der Kontrolle sei man denjenigen ausgeliefert, die nun Kontrolle haben über die eigenen Informationen: „(…) it is the increased risk of losing privacy that is one of the considerations that renders a loss of control over information (or space) problematic."[712] Denn

708 Thomson 1975:313
709 Ebd.:308
710 Macnish 2018:426ff.
711 Ebd.:429
712 Ebd.:423

durch den Kontrollverlust fühlt man sich nicht nur verletzlich, man ist auch verletzlich.[713]

Auch wenn in dieser Arbeit, mit Bezug auf Mohr, davon ausgegangen wird, dass sich das Recht auf Privatsphäre nur aus der Pflicht zum Schutz der Privatsphäre ableiten lässt[714], wird trotzdem kein reduktionistischer Ansatz vertreten. Denn das aus der Pflicht zum Schutz der Privatsphäre resultierende Recht auf Privatsphäre kann dennoch als eine eigene Art von Recht anerkannt werden.

Nicht-Reduktionismus

Parent argumentiert, es wäre vernünftiger und verständlicher, von einem Grundrecht auf Privatsphäre zu sprechen, das unter bestimmten Umständen Personen Schutz bieten kann, die nicht gesehen oder gehört werden wollen, oder die nicht wollen, dass ihr Eigentum beobachtet wird: „Privacy is a fundamental human value in societies like ours. It makes perfectly good sense for us, then, to acknowledge a fundamental right to privacy."[715] Auf diese Weise müssten Privatsphärenverletzungen nicht mehr mit Hilfe der Verletzung anderer Rechte erklärt werden.

Der Behauptung, dass das Recht auf Privatsphäre ein eigenständiges Recht sein soll, würden Manson und O'Neill zustimmen. Sie begründen das Recht auf Privatsphäre damit, dass die Menschen ein Interesse an ihrer Privatsphäre haben.[716] Sie gehen sogar einen Schritt weiter und postulieren, dass alle Menschen ein Interesse am Schutz der Privatsphäre haben sollten: „It seems uncontentious enough that each of us may have reason to keep certain information from becoming known by others, or by certain others, without our consent."[717] Aber Moore bemerkt die Schwäche dieses Argumentes – seine Subjektivität. Denn wenn die Interessen der privaten Menschen relevant sind, dann wären auch die Interessen der Privatsphäre-Verletzenden ausschlaggebend: „(…) preferences can be arbitrary

713 Macnish 2018:419, 430
714 Mohr 2010:77
715 Parent 1983b:350
716 Manson, O'Neill 2007:98
717 Ebd.:98f.

(...)."[718] Um diesem Widerspruch aus dem Weg zu gehen, argumentiert Moore, dass ein eigenständiges Recht auf Privatsphäre wertvoll ist, weil es für das menschliche Wohlbefinden unerlässlich ist und außerdem durch den Anspruch auf Nutzung und Kontrolle der eigenen Informationen keinen Schaden für jemand anderen entsteht.[719]

Laut Moore darf eine Person nicht beobachtet oder belauscht werden, weil dadurch Risiken entstehen, die für das Wohlbefinden der beobachteten oder belauschten Person moralisch relevant sind. Das Beobachten oder Belauschen selbst mag die Situation der betroffenen Person nicht direkt verschlechtern, aber eine solche Praxis des gegenseitigen Beobachtens und Belauschens würde die betroffenen Personen schlechter stellen. Laut Moore haben Individuen das Recht auf Privatsphäre, nicht weil ihr materielles Wohlergehen dadurch geschädigt wird oder weil sie den allgemeinen Wunsch haben, nicht beobachtet zu werden – dann wäre der Wunsch zu beobachten auch moralisch relevant. Sondern Individuen haben ein Recht auf Privatsphäre, weil ihnen durch die Verletzung der Privatsphäre ein unzulässiges Risiko auferlegt würde.[720] Da in dieser Arbeit davon ausgegangen wird, dass ein Recht auf Privatsphäre nur aus einer Pflicht zum Schutz der Privatsphäre resultieren kann[721], haben zuerst einmal Datengenerierende eine Pflicht zum Schutz der Privatsphäre von Datenquellen, weil Erstere den Letztgenannten durch die Verletzung der Privatsphäre ein unzulässiges Risiko auferlegen würden. Daraus kann dann ein Recht der Datenquellen auf ihre intakte Privatsphäre abgeleitet werden.

Es gibt aber auch Einwände gegen die Pflicht zum Schutz der Privatsphäre: So ließe sich argumentieren, dass die Risiken, die durch das Missachten der Privatsphäre entstehen können, ausgeglichen werden durch mehr Möglichkeiten und mehr Sicherheit.[722] Denn durch die Überwachung aller Menschen können die Böswilligen unter ihnen identifiziert und davon abgehalten werden, Schaden anzurichten. Aber die bei der Überwachung zusammengetragenen Informationen können auch für andere Zwecke als die Identifikati-

718 Moore 2010:34
719 Ebd.
720 Moore 2010:97
721 Mohr 2010:77
722 Ebd.:91

on von Terroristen genutzt werden. Diese Zweckentfremdung der Daten nennt sich data shifting. So könnten mit Hilfe der Terrorismus-Überwachungsdaten auch politische Einstellungen beobachtet und bewertet werden. Dies kann, genau wie beim chinesischen „Dangan", zu unterschiedlicher Behandlung von Menschen je nach politischer Einstellung führen. Dieses data shifting ist besonders deshalb gefährlich, weil sich die Risiken der multiplen Datennutzung erhöhen können. Wenn zum Beispiel die Regierung wechselt und sich als autokratisch erweist, entstehen neue Risiken, die zu Beginn nicht vorausgesehen wurden. Daraus lässt sich ein starkes Argument gegen diesen Einwand ableiten. O'Neill dreht das Argument sogar um: „Privacy is necessary for a degree of security (…)."[723] Man kann eine Einschränkung der Privatsphäre nicht mit einem Gewinn an Sicherheit rechtfertigen, weil der Schutz der Privatsphäre selbst wichtig ist, um Sicherheit zu ermöglichen. Denn beim Verletzen der Privatsphäre entstehen Risiken, welche die Sicherheit der Betroffenen gefährden.

Ein zweiter Einwand gegen die Pflicht zum Schutz der Privatsphäre ist, dass Individuen durch ihr Verhalten und das Teilen von Informationen in der Öffentlichkeit der Kontrolle über ihre Informationen durch andere zustimmen. Allerdings bedeutet das Verhalten in der Öffentlichkeit nicht, dass man einer zweckentfremdeten Nutzung oder einer Sammlung dieser Informationen zustimmt. Für viele Menschen ist es eine Notwendigkeit, in der Gesellschaft zu erscheinen – das ist nicht gleichzusetzen mit dem Einverständnis zur Überwachung ihres Verhaltens.[724] Hier ist wieder Moores Unterschied zwischen Zugang und Benutzung relevant: Nur weil eine Person den Zugang zu ihren Informationen gewährt, heißt das nicht, dass sie auch mit jeglicher Benutzung ihrer Informationen einverstanden ist.[725]

Die moralische Pflicht zum Schutz der Privatsphäre führt laut Mohr zu Legitimitäts-Forderungen an die Rechtsordnung: „Die Menschheit hat die moralische Pflicht, ein System von Rechten zu etablieren, welches sicherstellt, dass die Adressaten von Pflichten rechtlich zur Erfüllung ihrer Pflichten gezwungen werden können. Dazu sind moralische Rechte nicht imstande, nur [mit] juridischen

723 O'Neill 2021:80
724 Moore 2010:92
725 Ebd.:28

Rechten korrelieren Rechtspflichten, und nur diese sichern die Erfüllung von berechtigten Ansprüchen."[726] Ein Recht bleibt aber etwas Institutionelles, das aus einer moralischen Pflicht zum Schutz motiviert sein kann. Das ist in der Geschichte schon oft geschehen. Im Jahr 1890 zum Beispiel forderten die Juristen Samuel D. Warren und Louis D. Brandeis die Entwicklung von Gesetzen parallel zu der Neuerung, dass Fotografien von Privatpersonen in Zeitungen verbreitet wurden. Neue mechanische Geräte ermöglichten es plötzlich, Privates zu veröffentlichen: „For years there has been a feeling that the law must afford some remedy for the unauthorized circulation of portraits of private persons; and the evil of the invasion of privacy by the newspapers (…)."[727] Heute, 133 Jahre später, gibt es eine europäische Rechtsprechung, die den Schutz unserer Privatsphäre im Angesicht der neuen technischen Entwicklungen gewähren soll. Die rechtliche Entwicklung macht deutlich, dass Datengenerierende eine Pflicht zum Schutz der Privatsphäre haben, zu der sie mit Hilfe der Rechtsprechung zur Verantwortung gezogen werden können.

6.3.1 Aktueller rechtlicher Rahmen im Vergleich zu ethischen Anforderungen

In der Europäischen Union ist Datenschutz rechtlich verankert in der Datenschutzgrundverordnung (DSGVO): „Diese Verordnung enthält Vorschriften zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Verkehr solcher Daten."[728] Für die Verarbeitung von personenbezogenen Daten wurden folgende Grundsätze erstellt: Rechtmäßigkeit, Verarbeitung nach Treu und Glauben, Transparenz, Zweckbindung, Datenminimierung, Richtigkeit, Speicherbegrenzung, Integrität und Vertraulichkeit und Rechenschaftspflicht.[729] Die Verarbeitung personenbezogener Daten ist nach der DSGVO dann rechtmäßig, wenn die betroffenen Person ihre Einwilligung für einen oder mehrere bestimm-

726 Mohr 2010:77
727 Warren, Brandeis 1890:195
728 Artikel 1, Absatz 1 DSGVO
729 Artikel 5 DSGVO

te Zwecke gibt.[730] Die Erklärung zur Einwilligung muss „(...) in verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache (...)"[731] formuliert sein und die Einwilligung der betroffenen Person muss nachweisbar und jederzeit leicht widerrufbar sein.[732] Außerdem soll die Einwilligung freiwillig auf der Basis von Informationen über die Verarbeitung und ihre Zwecke erfolgen.[733] Dabei soll die Zustimmung *aktiv* erfolgen und nicht auf Grund einer vorausgewählten Einwilligung. Zudem kann ein Vertrag, eine rechtliche Verpflichtung zum Schutz lebenswichtiger Interessen von natürlichen Personen sein, und öffentliches Interesse und die Wahrung berechtigter Interessen von natürlichen Personen kann ein rechtmäßiger Grund für die Verarbeitung personenbezogener Daten sein.[734] Die Personen, deren Daten verarbeitet wurden, haben unter anderem ein Recht auf Auskunft, Berichtigung unrichtiger Daten und Löschung von falschen oder nicht mehr benötigten Daten.[735] „Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt."[736] Dabei muss es immer eine verantwortliche Person für die Datenverarbeitung geben[737], und personenbezogene Daten dürfen nicht beliebig weitergegeben werden[738].

Auch das deutsche Grundgesetz schützt die Kontrolle von Individuen über ihre Informationen. Das Bundesverfassungsgericht hat im Jahr 1983 geurteilt, dass sich aus der Menschenwürde (GG Artikel 1, Absatz 1) und dem allgemeinen Persönlichkeitsrecht (GG Artikel 2, Absatz 1) ein Recht auf informationelle Selbstbestimmung ableiten lässt: „Das Grundrecht gewährleistet insoweit die Befugnis des

730 Art. 6, Absatz 1 a) DSGVO
731 Art. 7 DSGVO
732 Ebd.
733 Erwägungsgrund 32 DSGVO
734 Art 6, Absatz 1 b)-f) DSGVO
735 Artikel 15–17 DSGVO
736 Artikel 22, Absatz 1 DSGVO
737 Artikel 24–43 DSGVO
738 Artikel 44–50 DSGVO

Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen."[739]

Bedrohungen der Privatsphäre entstehen dem Juristen Daniel Solove zufolge durch zum Beispiel Überwachung, Datensammlung, persönliche Identifizierung und Zweckentfremdung, sodass die Verletzung der Vertraulichkeit und persönliche Bloßsstellung, sowie der Ausschluss von Individuen, Einflussnahme auf Entscheidungen oder Erpressung möglich werden.[740] Die Ingenieurwissenschaftlerin Marie Caroline Oetzel und die Wirtschaftsinformatikerin Sarah Spiekermann haben im Jahr 2014 den damaligen Entwurf der DSGVO mit den möglichen Bedrohungen der Privatsphäre abgeglichen und kamen zu dem Schluss, dass die Datenschutzvorschriften die Eindämmung aller potenziellen Gefahren für die Privatsphäre unterstützen.[741]

Ich argumentiere jedoch, dass im Recht ausschließlich ein ethischer Mindeststandard festgehalten ist, der im Gegensatz zur Ethik auch Ausnahmen vorsieht. Laut Artikel sechs der Datenschutzgrundverordnung ist die Datenverarbeitung auch dann rechtmäßig, wenn die betroffene Person der Verarbeitung nicht zugestimmt hat. Denn weitere Bedingungen für die Rechtmäßigkeit der Datenverarbeitung sind die Erfüllung eines Vertrages (b), die Erfüllung einer rechtlichen Verpflichtung (c), der Schutz von lebenswichtigen Interessen von Personen (d), die Wahrnehmung einer Aufgabe, die im öffentlichen Interesse liegt oder in Ausübung öffentlicher Gewalt erfolgt (e), oder die Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten (f).[742] Dies gilt auch für das Übermitteln personenbezogener Daten an ein Drittland oder an eine internationale Organisation. So können auch hier Verträge (b und c), öffentliches Interesse (d), Rechtsansprüche (e) oder lebenswichtige Interessen (f) ausschlaggebend sein.[743]

Eigentlich sollten die betroffenen Personen ihre Einwilligung zu jedem konkreten Verarbeitungszweck ihrer Daten geben können[744],

739 BVerfG, Urteil v. 15. Dezember 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83
740 Solove 2006; Oetzel, Spiekermann 2014:130
741 Oetzel, Spiekermann 2014:130
742 Artikel 6 (1) DSGVO
743 Artikel 49 (1) DSGVO
744 Erwägungsgrund 32 DSGVO

aber auch hier gibt es Ausnahmen. Denn insbesondere im Kontext von wissenschaftlicher Forschung ist es schwierig, jeden einzelnen Verarbeitungszweck schon während der Datenerhebung anzugeben. Deshalb ist es laut Erwägungsgrund 33 der DSGVO möglich, nur eine breite Zustimmung für bestimmte Bereiche wissenschaftlicher Forschung einzuholen.[745] Ähnlich urteilte das Bundesverfassungsgericht, dass es auch bei der Datenerhebung für Statistiken nicht möglich sei, die Einwilligung für jeden konkreten Zweck einzuholen.[746] Zudem gilt laut DSGVO der Datenschutz nur für Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen. Das heißt, anonymisierte Informationen sind nicht durch die DSGVO geschützt, da die Kosten und der Zeitaufwand zur Identifizierung von natürlichen Personen unverhältnismäßig hoch wären.[747]

Die Ethik hingegen sieht keine Ausnahmen vor. Prinzipiell sollten Datenquellen immer die Kontrolle darüber haben, mit wem sie welche Informationen zu welchem Zweck teilen. Durch breite Zustimmung zu verschiedenen potenziellen Verarbeitungszwecken verlieren Menschen jedoch die Kontrolle über die Verwendung ihrer Daten. Wenn die Daten für unbekannte Zwecke verwendet werden, können die Datenquellen nicht mehr kontrollieren, mit wem sie welche Informationen zu welchem Zweck teilen. Laut DSGVO soll die Anonymisierung von Informationen ein wirksames Mittel zum Schutz der Privatsphäre sein.[748] Aus der ethischen Perspektive unterstützt das Anonymisieren von Daten den Schutz von Individuen allerdings nicht ausreichend, da nur die vermeintlich personenbezogenen Informationen gelöscht (oder bei Pseudonymisierung separat gespeichert)[749] werden. Jedoch müssen Informationen nicht personenbezogen sein, um durch ihre Verarbeitung eine Gefahr für ein Individuum darzustellen.[750] Denn es ist möglich, dass aus nichtpersönlichen Informationen, dann doch persönliche Informationen über geschützte Merkmale abgeleitet werden.[751] Ein hoher Kosten-

745 Erwägungsgrund 33 DSGVO
746 BVerfG, Urteil v. 15. Dezember 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83
747 Erwägungsgrund 26 DSGVO
748 Ebd.
749 Wimmer 2019
750 Martani, Hummel 2022:37
751 Ebd.:34; Manson, O'Neill 2007:104; Wachter, Mittelstadt 2019:564

und Zeitaufwand hält die Ethik nicht davon ab, diese Möglichkeit zu problematisieren.

Die DSGVO als aktueller rechtlicher Rahmen bildet die Mindestanforderungen der Ethik weitestgehend gut ab. Denn auch die ethische Erörterung der Privatsphäre in Bezug auf Datenverarbeitung ergibt, dass die Datenquellen geschützt werden sollten. Die DSGVO setzt das um, indem das Verhalten der Datengenerierenden reguliert wird. Die ethische Analyse sieht vor, dass Außenstehende bei Individuen keinen Kontrollverlust verursachen dürfen, indem sie bewusst versuchen, auf persönliche Informationen des Individuums zuzugreifen und diese zu benutzen oder tatsächlich zugreifen, auch aus Versehen, ohne dass das Individuum darüber die Kontrolle hat. Außerdem sollten Außenstehende nicht durch fahrlässiges Handeln Beihilfe zur Privatsphärenverletzung leisten, indem sie das Individuum dem substanziell größeren Risiko aussetzen, dass eine andere außenstehende Person auf seine persönlichen Informationen zugreift. Denn Datengenerierende haben eine moralische Pflicht zum Schutz der individuellen Privatsphäre. Deshalb ist es wichtig, dass die Handlungen und die Kommunikation, die zur Beschaffung von persönlichen Informationen, sowie für die Verwendungen von solchen Informationen genutzt werden, moralisch zulässig sind.[752] In der DSGVO geht es nicht vorrangig um die moralische Zulässigkeit der mit Datenverarbeitung verbundenen Handlungen und Kommunikation – es geht um die rechtliche Zulässigkeit. In Bereichen wie der Zweckgebundenheit der Verarbeitung oder dem Schutz von anonymisierten Daten wäre die Ethik strenger als das geltende Recht. Trotzdem bietet die DSGVO eine gute Grundlage für die ethische Regulierung von Datenverarbeitung. Denn durch diese Verordnung wurde die moralische Pflicht zum Schutz der Privatsphäre in juristisches Recht gegossen.

752 Manson, O'Neill 2007:110

6.4 Aus der Pflicht zum Schutz der Privatsphäre abgeleitete Anforderungen

In diesem Kapitel ist klar geworden, dass Datengenerierende zum Schutz der Privatsphäre der Datenquellen moralisch verpflichtet sind, und dass die Privatsphäre der Datenquellen aber auf Grund des geplanten oder tatsächlichen Zugriffs auf ihre Daten ohne ihre Wahl-Kontrolle verletzt werden kann. Die Verletzung der Privatsphäre kann auf Grund ihrer unerwünschten Folgen von ungewollten Eingriffen und fehlender Autonomie als Risikoauferlegung bezeichnet werden. Mit Hilfe der deontologischen Risikoethik lässt sich deshalb argumentieren, dass diese Risikoauferlegung nur im Rahmen eines gerechten sozialen Systems der Risikobereitschaft geschehen darf, welche allen Beteiligten zum Vorteil gereicht.[753] Das bedeutet, durch die Regel der dritten Formel des Kategorischen Imperativs von Kant, dass niemand als bloßes Mittel gebraucht werden darf.[754] Das heißt in der Praxis, die individuelle Kontrolle der Datenquellen über das Risiko sollte so groß wie möglich sein. Dafür sollten die Risikoexponierten ohne Täuschung über das Risiko informiert werden und ohne Zwang die Wahl haben, ob sie das Risiko tragen wollen. Auf diese Weise kann die Privatsphäre sowohl geschützt als auch gefördert werden.

Aus der Definition von Privatsphäre als Wahl-Kontrolle des Individuums über Zugang zu und Benutzung von seinen persönlichen Informationen lassen sich Anforderungen an den Umgang mit Daten und insbesondere mit den Datenquellen im Datenökosystem ableiten. Wie in Tabelle 3 zu sehen ist, drückt die Definition zwei Ziele aus: Ein Individuum sollte die Wahl-Kontrolle sowohl über den Zugang zu als auch die Benutzung von seinen persönlichen Informationen haben. Da Wahl-Kontrolle bedeutet, dass man entscheiden kann, mit wem man welche Informationen zu welchem Zweck teilt, sollten die betroffenen Personen also einerseits entscheiden können, wer Zugriff auf welche Informationen bekommt. Andererseits sollten die betroffenen Individuen entscheiden können, zu welchem Zweck ihre Informationen gesammelt oder verarbeitet werden. Administrativ bedeutet das, dass einerseits das Einverständnis der Individuen

753 Hansson 2003:305
754 Kant 2008: 65 [429]

eingeholt werden sollte, und die Individuen andererseits über jeden Verwendungszweck der Daten informiert sowie die Daten vor Zweckentfremdung geschützt werden sollten.

Tabelle 3 Administrative Anforderungen für die Umsetzung von Privatsphäre[755]

Privatsphäre Definition	**Privatsphäre Ziele**	**Umsetzung des Ziels**	**Administrative Anforderungen**
Privatsphäre = Wahl-Kontrolle des Individuums über Zugang zu und Benutzung von seinen persönlichen Informationen	Das Individuum hat die Wahl-Kontrolle über den *Zugang* zu seinen persönlichen Informationen	Das Individuum kann entscheiden, wer *Zugriff* auf welche Informationen bekommt	*Einverständnis* der Individuen wird eingeholt
			Gewährleistung von *Freiwilligkeit* der Einwilligung
	Das Individuum hat die Wahl-Kontrolle über die *Benutzung* von seinen persönlichen Informationen	Das Individuum kann entscheiden, zu welchem *Zweck* seine Informationen gesammelt oder verarbeitet werden	Individuen werden über alle Verwendungszwecke *informiert*
			Gewährleistung von *Zweckentsprechung* der Informationen

Diese administrativen Anforderungen sorgen dafür, dass die Datenquellen in Bezug auf den Umgang mit ihren Daten im Datenökosystem selbstbestimmt handeln können und sie somit nicht als bloßes Mittel gebraucht werden. Da die Auferlegung eines Risikos nur dann zulässig ist, wenn niemand als bloßes Mittel gebraucht wird, ist die Erfüllung der administrativen Anforderungen eine notwendige Bedingung für die Zulässigkeit der Risikoauferlegung.

Aus dem Prinzip der Privatsphäre lässt sich also die Anforderung ableiten, dass Individuen sowohl nach ihrem Einverständnis gefragt werden, bevor ihre Informationen gesammelt und gespeichert werden als auch, dass Individuen vorher darüber informiert werden, zu welchem Zweck ihre Informationen gesammelt, gespeichert und verarbeitet werden. Die in der DSGVO verankerte Anforderung der informierten Einwilligung lässt sich also direkt aus der Definition von Privatsphäre als Wahl-Kontrolle über Zugang zu und Benutzung

755 Eigene Darstellung

von persönlichen Informationen ableiten. Während aber gemäß der DSGVO auch Verträge, öffentliches Interesse oder lebenswichtige Interessen rechtmäßige Gründe zur Datenverarbeitung sein können, handelt es sich aus ethischer Sicht immer dann um eine Verletzung der Privatsphäre, wenn Außenstehende auf persönliche Informationen des Individuums zugreifen, ohne dass das Individuum darüber die Wahl-Kontrolle hat.

In einem Datenökosystem sind Datenquellen dem Risiko ausgesetzt, dass ihre Privatsphäre verletzt wird, weil ihnen oft die Kontrolle über den Zugang zu und die Benutzung ihrer persönlichen Informationen fehlt. Diese Situation wird durch epistemische Asymmetrie erzeugt, da die Datenquellen in der Realität oft nicht wissen, wer zu welchen Zwecken ihre Informationen sammelt und auswertet. Wenn zum Beispiel ein Supermarkt das Kaufverhalten seiner Kundinnen beobachtet, um herauszufinden, welche von ihnen schwanger ist, ohne die Kundinnen zu informieren und um ihr Einverständnis zu bitten, dann fehlt den Kundinnen das Wissen darüber, dass ihre Privatsphäre verletzt wird. Aber, selbst wenn die Kundinnen informiert sind, können die Informationen über ihr Kaufverhalten trotzdem ohne ihr Wissen zweckentfremdet werden. Durch dieses fehlende Wissen über den Umgang von Außenstehenden mit den eigenen Daten ist es wahrscheinlicher, dass die Privatsphäre von Datenquellen tatsächlich verletzt wird. Denn die Privatsphäre als Wahl-Kontrolle über Zugang zu und Benutzung von persönlichen Informationen wird verletzt, wenn die Datenquellen gar nicht wissen, dass Zugang zu und Benutzung von ihren Informationen stattfindet.

Das fehlende Wissen und die dadurch mögliche Verletzung der Privatsphäre werden durch die Struktur eines Datenökosystems begünstigt. Denn Datengenerierende sind zu einem gewissen Grad dazu in der Lage, die für Datenquellen verfügbaren Beweise in Bezug darauf, was mit den Daten passiert, zu kontrollieren – das ermöglicht epistemische Dominanz.[756] Epistemische Dominanz kann bösartig sein, wenn der dominierte Akteur absichtlich daran gehindert wird, epistemisch wertvolle Zustände zu erreichen. Aber wenn die dominierenden Akteure wohlmeinend und gut informiert sind, kann

756 Harris 2022:134, 137

das den dominierten Akteuren epistemisch zugutekommen.[757] Aus der Pflicht zum Schutz der Privatsphäre entsteht auch eine Pflicht des dominierenden Akteurs, wohlmeinend zu sein. Wenn die Privatsphäre nicht geschützt wird oder sogar durch bösartige epistemische Dominanz verletzt wird, besteht das Risiko, dass die betroffenen Individuen neben dem Kontrollverlust noch weitere Schäden erleiden durch zum Beispiel Diskriminierung oder eine Einschränkung ihrer Autonomie. Angenommen, wir wollen in einer Welt leben, in der sich die Menschen nicht gegenseitig Schaden zufügen, so sollte epistemische Dominanz nicht bösartig ausgenutzt werden, um möglichst die Verletzung von Privatsphäre zu vermeiden.

Mainz und Uhrenfeldt sind der Meinung, dass die Verletzung der Privatsphäre aus gewichtigeren Gründen, die nichts mit der Privatsphäre zu tun haben, gerechtfertigt sein kann.[758] Da die Datenquellen bei der Verletzung ihrer Privatsphäre jedoch Risken ausgesetzt werden, ist eine solche Verletzung nur unter den Bedingungen der deontologischen Risikoethik zulässig – unabhängig davon, wie gewichtig die Gründe sind. Solange die administrativen Anforderungen eingehalten werden, kann die Risikoexposition der Datenquellen im Datenökosystem, in Form von der Verletzung ihrer Privatsphäre, Teil eines gerechten sozialen Systems der Risikobereitschaft sein. Als nächstes ist es wichtig, auch das Risiko zu betrachten, dem die Datennutzenden im Datenökosystem ausgesetzt sind.

757 Harris 2022:136f., 139

758 Mainz, Uhrenfeldt 2021:295, Fußnote 18

7 Transparenz

In diesem Kapitel wird erarbeitet, wieso das Fehlen von Transparenz überhaupt ein unerwünschtes Ereignis wäre, warum dieses Fehlen eine verbotene Handlungsfolge wäre und wie man die Transparenz schützen bzw. fördern kann. Dafür ist es zuerst einmal notwendig zu definieren, was genau mit Transparenz im Kontext von Datenökosystemen gemeint ist – insbesondere wie Transparenz im Datenökosystem hergestellt werden kann. Ich zeige auf, dass das Vorgehen des Algorithmus bekannt sein muss, damit seine Ergebnisse auf Verzerrungen überprüft werden können. Zu diesem Zweck können Post-hoc Erklärungen von Algorithmen das Verhältnis zwischen Input und Output offengelegen und dadurch aufzeigen, warum der Algorithmus eine bestimmte Ausgabe erzeugt. Damit dieser Zusammenhang verständlich und das Ergebnis damit transparent ist, schlage ich vor, dass die Ergebnisse im zum Verständnis notwendigen Kontext dargestellt werden sollten. Die moralische Pflicht zur transparenten Darstellung algorithmischer Ergebnisse durch Datenverarbeitende leite ich aus Kants erster Formulierung des Kategorischen Imperativs ab. Daraus resultiert für Datennutzende ein Recht auf Transparenz, welches sich zum Teil schon in europäischen und deutschen Gesetzen wiederfinden lässt. Abschließend lassen sich aus der Pflicht zur Transparenz zwei administrative Anforderungen ableiten, welche über das bestehende Recht hinausgehen.

Transparenz ist das Gegenteil von Undurchsichtigkeit (opacity) und wird im Folgenden, wie bei Mittelstadt et al., als Zugänglichkeit und Verständlichkeit verstanden.[759] Bei Datennutzenden entsteht ein Risiko durch die fehlende Transparenz der algorithmischen Ergebnisse, weil es bei fehlender Transparenz möglich ist, dass den Datennutzenden verzerrte Ergebnisse zur Verfügung gestellt werden, auf deren Grundlage sie zu falschen Überzeugungen gelangen. Transparenz fehlt deshalb, weil Datennutzenden sowohl der Zugang zu als

759 Mittelstadt et al. 2016:6

auch das Verständnis von Informationen über die Entstehung der Ergebnisse fehlt.[760] Dies ist möglich auf Grund der epistemischen Asymmetrie im Datenökosystem. Denn Datenverarbeitende sind zu einem gewissen Grad dazu in der Lage, die für Datennutzende verfügbaren Beweise in Bezug darauf, wie das Ergebnis zustande gekommen ist, zu kontrollieren. Es handelt sich dabei nicht um eine vollkommene epistemische Dominanz, aber es besteht eine gewisse Abhängigkeit der Datennutzenden von den Datenverarbeitenden.[761] Auf Grund dieser Abhängigkeit besteht bei den Datennutzenden ein Interesse an Transparenz.

Damit die Datennutzenden die ihnen zur Verfügung gestellten Ergebnisse auf ihre Richtigkeit prüfen können und so möglichst nicht zu falschen Überzeugungen gelangen, benötigen sie Zugang zu und Verständnis von Informationen über die Entstehung der Ergebnisse. Diese Informationen können nur durch die Datenverarbeitenden zur Verfügung gestellt werden, aber auch die kennen auf Grund von der Nutzung komplizierter Algorithmen nicht unbedingt den gesamten Entstehungsprozess. In Bezug auf algorithmische Ergebnisse bedeutet Transparenz also, dass man den Mechanismus, nach dem der eingesetzte Algorithmus funktioniert, in gewisser Weise versteht.[762] Es ist also unklar, wie Transparenz im Datenökosystem hergestellt werden kann, es ist jedoch klar, dass Transparenz im Datenökosystem wichtig ist, um falsche Ergebnisse und darauf gegründete falsche Überzeugungen vermeiden zu können – dieses systematische Problem soll in diesem Kapitel vertieft und gelöst werden.

Da in Datenökosystemen viele Daten verarbeitet werden, nutzen Datenverarbeitende dafür oft Hilfsmittel in Form von Algorithmen. Hier gilt es zu unterschieden zwischen Wenn-Dann-Algorithmen und Black Box KI-Algorithmen (auch KI-Systeme genannt). Wenn-Dann-Algorithmen können als „White Box" bezeichnet werden, weil es sich um einen vorgegebenen Satz von Anweisungen handelt.[763] Solche Algorithmen sind transparent, da deren Ergebnisse problemlos zugänglich gemacht und verständlich dargestellt werden können. KI-Systeme hingegen sind nicht deterministisch – wenn sich ihre Ei-

760 Mittelstadt et al. 2016:6
761 Harris 2022:134, 137
762 Lipton 2018:42
763 Tutt 2017:107

genschaften dennoch leicht vorhersagen und erklären lassen, nennt man das „Grey Box". Sobald Algorithmen emergente Eigenschaften aufweisen, die es schwierig oder unmöglich machen, ihre Merkmale vorherzusagen oder zu erklären, handelt es sich um eine „Black Box".[764] So ein Black Box Algorithmus verbirgt Annahmen und Vereinfachungen. Aufgrund dieser Undurchsichtigkeit können Datennutzende die in den Algorithmus eingebettete Weltanschauung nicht erkennen und seine Glaubwürdigkeit nicht überprüfen. Black Box KI-Systeme sind also nicht inhärent transparent, da es schwierig bis unmöglich ist, die Entstehung ihrer Ergebnisse verständlich darzustellen. Da es darum geht herauszufinden, wie die Entstehung der Ergebnisse von KI-Systemen dennoch transparent dargestellt werden kann, ist im Folgenden mit Algorithmus immer ein Black Box KI-Algorithmus gemeint.

Das größte Problem der fehlenden Transparenz beim Einsatz von Algorithmen ist, dass ihre Ergebnisse, auch bezeichnet als Entscheidungen, nur mit einer gewissen Wahrscheinlichkeit richtig sind. Sie können falsch sein oder anders, als ein Mensch sie getroffen hätte. Insbesondere, wenn auf Grundlage einer solchen Entscheidung Klassifizierungen vorgenommen werden, ist eine andere Entscheidung, als ein Mensch sie getroffen hätte, nicht notwendigerweise schlecht. Denn Menschen haben Vorurteile, und es fällt ihnen oft schwer, objektive Entscheidungen zu fällen. Die Hoffnung wäre, dass Algorithmen objektiv entscheiden und damit Ungerechtigkeiten vermeiden. Aber das stimmt leider nicht, denn auch algorithmische Ergebnisse sind häufig von menschlichen Vorurteilen beeinflusst. Einerseits, weil Menschen zum Beispiel die Trainingsdaten vorklassifizieren sowie Schwellenwerte und Parameter anpassen und andererseits, weil die Trainingsdaten immer aus menschlichen Kontexten entstehen und deshalb auch menschliche Vorurteile widerspiegeln.[765] Auf diese Weise kann das Ergebnis auf Entscheidungskriterien basieren, die Verzerrungen aus bisherigen Entscheidungen reproduzieren, zum Beispiel bei dem Algorithmus, der von Amazon im Rekrutierungsprozess eingesetzt wurde. In der Vergangenheit waren es vor allem Männer, die in Amazons Bewerbungsverfahren erfolgreich waren. Dieses Muster hatte zur Folge, dass ein Entschei-

764 Tutt 2017:107
765 Burrell 2016:3

dungskriterium des Algorithmus das Geschlecht der Bewerbenden war. Deshalb wurden Bewerbungen von Frauen vom KI-System abgelehnt – selbst wenn das Geschlecht nicht direkt genannt wurde, sondern nur z.B. durch feminin konnotierte Hobbys erkennbar war.[766] Das KI-System wurde bei Amazon zum Identifizieren der besten Bewerbungen eingesetzt, bis anhand der Ergebnisse eine Verzerrung deutlich wurde.

Solch eine Verzerrung nennt sich im Englischen „Bias", sie kommt zustande, wenn bestehende Diskriminierungsstrukturen Eingang in die Trainingsdaten eines Algorithmus finden. So schreiben der Technologiephilosoph Bryce Goodman und der Informatiker Seth Flaxman: „Indeed, machine learning can reify existing patterns of discrimination — if they are found in the training data set, then by design an accurate classifier will reproduce them."[767] Wird der Begriff „Bias" im Kontext mit Computersystemen verwendet, dann ist laut der bereits erwähnten Informationswissenschaftlerin Batya Friedman und dem Psychologen Peter H. Kahn Jr. gemeint, dass diese Computersysteme systematisch bestimmte Personen oder Personengruppen zugunsten anderer diskriminieren: „Bias refers to systematic unfairness perpetrated on individuals or groups."[768] Eine solche Diskriminierung ist unfair, wenn einer Person oder einer Gruppe von Personen aus unvernünftigen oder unangemessenen Gründen eine Chance oder ein Gut vorenthalten oder ihnen ein unerwünschtes Ergebnis zugewiesen wird. Eine unfaire Diskriminierung führt aber erst dann zu einem „Bias", wenn sie systematisch erfolgt und mit einem ungerechten Ergebnis einhergeht.[769]

Friedman und Nissenbaum unterscheiden drei verschiedene Formen von Verzerrungen: Bestehende, technische und emergente Verzerrungen. Bestehende Verzerrungen (preexisting bias) existieren unabhängig und in der Regel schon vor der Entwicklung des Systems. Bereits bestehende Vorurteile können in ein System einfließen, entweder durch ausdrückliche und bewusste Bemühungen von Einzelpersonen oder Institutionen oder implizit und unbewusst, sogar

766 Stahl et al. 2023:10f.
767 Goodman, Flaxman 2017:53
768 Friedman, Kahn 2008:1253
769 Friedman, Nissenbaum 1996:332

trotz der besten Absichten.[770] Es können sich sowohl gesellschaftliche Strukturen als auch individuelle Voreingenommenheit von Personen in Computersystemen widerspiegeln. Auf der gesellschaftlichen Ebene beeinflussen Organisationen, Institutionen oder die Kultur die Entstehung von Verzerrungen, während es auf der individuellen Ebene auf die Einstellungen von Personen ankommt, die einen wesentlichen Einfluss auf die Gestaltung des Systems haben.

Technische Verzerrungen (technical bias) ergeben sich aus technischen Zwängen oder Überlegungen. So kann eine technische Verzerrung zum Beispiel durch die Beschränkungen von Computerhardware und -software entstehen. Eine solche technische Beschränkung ist zum Beispiel die begrenzte Größe des Bildschirms, wodurch auf einer Website mit Flugangeboten nur einige Angebote direkt sichtbar sein können – die weiteren Angebote befinden sich auf der nächsten Seite. Wenn der Ranking-Algorithmus systematisch Flüge ausgesuchter Fluggesellschaften zuoberst und Flüge anderer Fluggesellschaften später platziert, weist das System eine technische Verzerrung auf. Technische Verzerrungen können ebenfalls entstehen, weil ein Algorithmus angewendet wird, der nicht alle Gruppen unter allen wichtigen Bedingungen gleichbehandelt oder weil Pseudozufallszahlen unzulänglich generiert werden, sodass eine bestimmte Gruppe von Menschen systematisch bevorzugt wird, oder weil der Versuch unternommen wird, menschliche Konstrukte wie Diskurse, Urteile oder Intuitionen für Computer zugänglich zu machen.[771] Darunter fällt zum Beispiel auch die von Goodman identifizierte Unsicherheitsverzerrung (uncertainty bias). Diese entsteht, wenn eine gesellschaftliche Gruppe im Datensatz unterrepräsentiert ist und der Algorithmus gleichzeitig risikoavers entscheidet. Denn dann bevorzugt der Algorithmus zum Beispiel bei einer Kreditvergabe die Gruppen, die in den Trainingsdaten besser repräsentiert sind, da mit diesen Vorhersagen weniger Unsicherheit verbunden ist.[772]

Emergente Verzerrungen (emergent bias) entstehen erst bei der Benutzung des Computersystems. Wenn zum Beispiel neue wissenschaftliche Erkenntnisse nicht in das Computersystem aufgenommen werden, entsteht eine emergente Verzerrung. Auch wenn sich

770 Friedman, Kahn 2008:1253; Friedman, Nissenbaum 1996:332–335
771 Friedman, Nissenbaum 1996:332, 334–336; Friedman, Kahn 2008:1253
772 Goodman 2016:5; Goodman, Flaxman 2017:54

die Bevölkerungsgruppe, die das System nutzt, von der Gruppe unterscheidet, die bei der Systemgestaltung als Nutzende angenommen wurde, entsteht eine emergente Verzerrung. So kann ein System zum Beispiel eine Lese- und Schreibfähigkeit voraussetzen, aber von Menschen angewandt werden, welche diese Fähigkeit nicht besitzen.[773]

Alle Verzerrungen können durch besondere Aufmerksamkeit minimiert oder sogar vermieden werden. Bestehende Verzerrungen (preexisting bias) können laut Friedman und Nissenbaum durch ein gutes Verständnis der herrschenden Vorurteile minimiert werden. Das heißt, bei der Gestaltung eines Computersystems müsste darauf geachtet werden, dass keine Voreingenommenheit in Bezug auf kulturelle Identität, Klasse, Geschlecht, Lese- und Schreibfähigkeit und Behinderung reproduziert werden. Hierfür könne es nützlich sein, eine diverse Gruppe von potenziell Nutzenden einzubeziehen, um das System zu testen und bewerten. So können auch unbeabsichtigte Verzerrungen aufgedeckt werden.[774]

Um technische Verzerrungen (technical bias) zu vermeiden, muss laut Friedman und Nissenbaum bei der Gestaltung schon die Nutzung des Systems antizipiert werden, sodass technische Entscheidungen nicht im Widerspruch zu moralischen Werten stehen. Bei der technischen Umsetzung eines Systems sollte es also nicht darum gehen, immer die einfachste Lösung zu realisieren, sondern auch darum, gerechte Ergebnisse zu produzieren.

Emergente Verzerrungen (emergent bias) können laut Friedman und Nissenbaum minimiert werden, indem antizipiert wird, in welchen Kontexten das System genutzt werden könnte. Wenn es für bestimmte Nutzungskontexte nicht möglich ist, das System passend zu gestalten, sollte das System in diesen Kontexten nicht angewendet werden.[775]

Laut Friedman und Nissenbaum liegt also die Verantwortung für die Vermeidung oder Minimierung von Verzerrungen bei denjenigen, die das System gestalten. Aber von den Verzerrungen betroffen sind vorrangig diejenigen, die das System nutzen und sich auf die Ergebnisse verlassen. Diese Gruppe der Datennutzenden ist nur

773 Friedman, Nissenbaum 1996:332, 335–336; Friedman, Kahn 2008:1253
774 Friedman, Kahn 2008:1253; Friedman, Nissenbaum 1996:343f.
775 Friedman, Nissenbaum 1996:344; Friedman, Kahn 2008:1253

dann in der Lage zu überprüfen, ob eine Verzerrung durch das System reproduziert oder ausgelöst wird, wenn die Vorgänge im System transparent sind. Transparenz setzt sich laut Mittelstadt et al. aus zwei Komponenten zusammen: Zugänglichkeit und Verständlichkeit.[776] Ohne Transparenz, also ohne zugängliche und verständliche Informationen über das System und sein Vorgehen, können Datennutzende nicht überprüfen, ob die Entscheidung des Systems auf Verzerrungen beruht und dadurch zusätzlich ein normatives Risiko für sie darstellt. Denn auf Grund ihrer Komplexität ist es nahezu unmöglich, im Voraus zu wissen, wann und wie Algorithmen versagen werden.[777]

Ob Computersysteme systematisch und unfair bestimmte Personen oder Personengruppen zugunsten anderer diskriminieren, lässt sich manchmal schon bei der Anwendung erkennen, wenn zum Beispiel die gute Sehfähigkeit der Nutzenden vorausgesetzt wird durch kleine Schrift oder geringe Kontraste. Andere Verzerrungen ließen sich auch beim Vergleich von Ergebnissen erkennen, wenn zum Beispiel Frauen systematisch und aus unvernünftigen oder unangemessenen Gründen andere Ergebnisse zur Verfügung gestellt bekommen als Männer und dies zu einem ungerechten Resultat führt. Allerdings ist es aufwendig, eine große Menge an Ergebnissen zu vergleichen, um eine Systematik feststellen zu können. Einfacher wäre es, wenn transparent wäre, mit Hilfe welcher Entscheidungskriterien das Computersystem zu den Ergebnissen kommt.

7.1 Umsetzung von Transparenz im Datenökosystem

In Bezug auf jegliche Algorithmen bedeutet Transparenz, dass zugänglich und verständlich sein muss, wie das Ergebnis zustande gekommen ist. Dies scheint jedoch schwer zu realisieren zu sein, da das Vorgehen von komplexen Algorithmen undurchsichtig ist.[778] Der Gedanke findet sich auch bei der KI-Ethikerin Kate Vredenburgh: „Some types of algorithms, such as decision trees, are usually judged by experts to be functionally transparent. Other algorithms, such

776 Mittelstadt et al. 2016:6
777 Tutt 2017:86
778 Burrell 2016:1; Humphreys 2009:619

as neural nets, are so complex that it is difficult to understand the rules by which inputs are related to outputs."[779] Durch die Komplexität der Algorithmen ist der Entstehungsprozess der Ergebnisse nicht transparent – er ist sogar undurchsichtig. Wie bereits unter 3.1 erwähnt, ist dem Epistemologen Paul Humphreys zufolge ein Prozess für eine Person epistemisch undurchsichtig (epistemically opaque), wenn die Person nicht alle epistemisch relevanten Elemente des Prozesses kennt.[780] Für Datennutzenden wäre ein epistemisch relevantes Element des Prozesses im Datenökosystem zu wissen, wie die Ergebnisse, die ihnen zur Verfügung gestellt werden, zustande gekommen sind.

Bei der Realisierung dieses Anspruchs gibt es zwei Probleme. Das eine ist das einseitige Interesse an Transparenz. Während Datennutzende ein Interesse daran haben, haben Datenverarbeitende eher ein Interesse an wirtschaftlicher Rentabilität. Dazu schreiben Mittelstadt et al.: „Information about the functionality of algorithms is often intentionally poorly accessible."[781] Der Zugang wird oft deshalb erschwert, da die Algorithmen den Unternehmen einen Wettbewerbsvorteil verschaffen. Eine vollständige Offenlegung der Algorithmen würde den Wettbewerb zwischen verschiedenen Anbietern beenden und wäre dementsprechend auch zum Nachteil der Nutzenden, weil ein Wettbewerb die Qualität verbessern kann. Außerdem würde die vollständige Transparenz des Algorithmus diesen womöglich mehr böswilligen Angriffen und Manipulationsversuchen aussetzen. Denn wer das System versteht, kann es zu seinem eigenen Vorteil ausnutzen.[782] Transparenz kann also auch Nachteile haben, aber trotzdem sind Datennutzende daran interessiert zu verstehen, wie die ihnen zur Verfügung gestellten Ergebnisse erstellt werden und wie sie Entscheidungen beeinflussen, die in datengesteuerten Verfahren getroffen werden.[783]

Außerdem ermöglicht Transparenz den Datennutzenden auch, mögliche wissentliche Täuschung, wie zum Beispiel die Umgehung von Vorschriften oder die Manipulation der Verbraucher durch Da-

779 Vredenburgh 2022:221
780 Humphreys 2009:618
781 Mittelstadt et al. 2016:6
782 Granka 2010:366; Kitchin 2016:7; Burrell 2016:3
783 Mittelstadt et al. 2016:6

tenverarbeitende, aufzudecken.[784] Bei der Abwägung, welche Interessen überwiegen – der Wettbewerbsvorteil durch Intransparenz oder mehr Kontrolle über datengesteuerte Ergebnisse durch Transparenz – haben Datenverarbeitende Vorteile, weil sie über die gefragten Informationen verfügen und sie dementsprechend vorenthalten können. Laut der Informatikerin und Soziologin Jenna Burrell bräuchte man unabhängige, vertrauenswürdige Prüfende, welche die Geheimhaltung für die Unternehmen wahren können und gleichzeitig dem öffentlichen Interesse dienen, indem sie Verbrauchermanipulationen oder Verstöße gegen Rechtsvorschriften durch Lesen des Codes erkennen.[785]

Das andere Problem ist, dass die Entscheidungsfindung von Algorithmen oft nicht eindeutig nachvollziehbar ist. Hier ergibt sich die Undurchsichtigkeit von Algorithmen für Datennutzende schon dadurch, dass nicht jede Person die Fähigkeit hat, Code zu schreiben und zu lesen.[786] Zudem werden die Methoden zur Programmierung von KI-Systemen immer komplexer, bis ihr Vorgehen selbst den Entwickler:innen nicht mehr verständlich ist. Der Rechtswissenschaftler Frank Pasquale formuliert das so: „Deconstructing the black boxes of Big Data isn't easy."[787] So ist auch das Identifizieren von Fehlern und ihren Ursachen erschwert.[788] Und es passieren laut den Technologiephilosoph:innen Thomas Grote, Konstantin Genin und Emily Sullivan tatsächlich Fehler beim Einsatz von Algorithmen: „Machine learning models often achieve impressive accuracy under training conditions, but fail in spectacular or unexpected ways when they are deployed in real-world settings."[789] Auch wenn Algorithmen im Test exakt arbeiten, können sie beim Einsatz in der Praxis versagen.

Das Vorgehen der Algorithmen ist so schwer zu verstehen auf Grund ihrer Undurchsichtigkeit (opacity). Algorithmen sind insofern undurchsichtig, als man als Empfänger der Ausgabe des Algorithmus keine konkrete Vorstellung davon hat, wie oder warum

784 Pasquale 2015:2; Burrell 2016:1f.
785 Burrell 2016:1f.
786 Ebd.
787 Pasquale 2015:6
788 Ertel 2021:343
789 Grote et al. 2024:1

die Ausgabe auf der Grundlage der Eingaben zustande gekommen ist.[790] Um dem entgegen zu wirken, könnte man fordern, dass die Öffentlichkeit im Verstehen von Code geschult wird, damit sie Algorithmen bewerten und kritisieren kann. Allerdings ist Burrell der Meinung, dass dabei die Komplexität der Algorithmen unterschätzt wird. Die Datensätze können überschaubar und der Code klar geschrieben sein, aber ihr Zusammenspiel im Algorithmus macht die Komplexität aus und sorgt dafür, dass es zeitaufwändig ist, ihn zu verstehen.[791] Die Berechnungen der Algorithmen sind so schnell und komplex, dass Menschen die Prozesse nicht reproduzieren oder verstehen können.[792] Selbst, wenn die Methoden öffentlich zugänglich wären, wäre es trotzdem noch schwer, sie zu verstehen.[793] Denn Algorithmen können für Personen undurchsichtig sein, weil eine Diskrepanz besteht zwischen der komplexen mathematischen Optimierung von Algorithmen und menschlichem Denken.[794]

Datennutzende sind typischerweise epistemisch von Datenverarbeitenden und den von ihnen eingesetzten Algorithmen abhängig. Die epistemische Abhängigkeit von algorithmischen Ergebnissen ist dem Technologieethiker Jeroen van den Hoven zufolge vergleichbar mit der epistemischen Abhängigkeit von Expert:innen. Denn es ist nicht möglich, alles aus eigener Erfahrung selbst zu wissen und jede Behauptung auf ihren Wahrheitsgehalt hin zu überprüfen – deshalb verlassen wir uns auf Expert:innen: „Many things we ordinarily claim to know, we claim to know because we believe that others possess direct and indefeasible (empirical) evidence for them."[795] Zum Beispiel behaupten wir zu wissen, dass Rauchen Lungenkrebs verursacht, obwohl wir das selbst nicht beweisen können. Um aber trotzdem Stellung nehmen zu können, verlassen wir uns auf Expert:innen. Die Rolle der Expert:innen übernehmen nun immer öfter Algorithmen, wenn sie zum Beispiel im medizinischen Kontext eingesetzt werden oder auch wenn sie entscheiden, ob eine Bewerberin abgelehnt wird. Datennutzende sind dann epistemisch von

790 Burrell 2016:1
791 Ebd.:4f.
792 Humphreys 2009:619
793 Pasquale 2015:6
794 Burrell 2016:1f.
795 Van den Hoven 1998:105

algorithmischen Ergebnissen abhängig, wenn es für sie schwierig ist, die Ergebnisse unabhängig zu verifizieren. Da die Schwierigkeit einer unabhängigen Verifizierung variieren kann, gibt es verschiedene Grade der epistemischen Abhängigkeit.[796]

Wenn Datennutzende sowohl eng eingebettet sind in so ein Datenökosystem als auch epistemisch abhängig von den algorithmischen Ergebnissen sind, bezeichnet van den Hoven sie als „epistemische Sklaven" des Systems, da es schwierig ist, dem System Output ohne zusätzliche Informationen zu widersprechen: „(…) if the system tells you that a candidate is unsuitable and has a record of high accuracy, it will be difficult to disagree (and defend the disagreement later on) if little other information is available."[797] Hier wenden Grote, et al. ein, dass die Vorhersagen des Modells mit großer Wahrscheinlichkeit nicht die einzigen Beweise sind, die den menschlichen Entscheidungstragenden zur Verfügung stehen, und sie sich unabhängig vom Algorithmus ein Urteil bilden können.[798] Van den Hoven würde jedoch widersprechen, denn seiner Meinung nach verwenden Datennutzende den Output der Algorithmen im Datenökosystem als direkten Input für ihre praktischen Überlegungen und stützen ihre Entscheidungen hauptsächlich darauf. Dadurch können Datennutzende für ihr eigenes Handeln womöglich keine vom Algorithmus unabhängigen Gründe mehr vorbringen, wenn zum Beispiel Bewerberinnen abgelehnt werden, weil der Algorithmus das empfiehlt.[799]

Auch bei menschlichen Empfehlungen kann man nicht davon ausgehen, dass sie frei von Verzerrungen sind. Es wäre also ein sehr hoher und womöglich unerfüllbarer Anspruch zu verlangen, dass man als Empfänger:in einer Empfehlung ausschließlich mit verzerrungsfreien und wahren Ergebnissen konfrontiert wird. In der Interaktion mit Menschen können wir nicht davon ausgehen, dass ihre Empfehlungen unbedingt wahr sind. Deshalb stellen wir Rückfragen oder prüfen die Empfehlungen mit Hilfe von anderen Quellen. Der Unterschied ist jedoch, dass Datennutzende bezüglich

796 Rooksby 2009:82; van den Hoven 1998:104f.; Buijsman, Veluwenkamp 2023:546
797 Buijsman, Veluwenkamp 2023:545
798 Grote et al. 2024:5
799 Van den Hoven 1998:104f.; Buijsman, Veluwenkamp 2023:543

der algorithmischen Ergebnisse oft keine Rückfragen stellen können (wenn es sich nicht um Chatbots handelt), und es zudem schwierig sein kann, alternative Quellen zu finden.[800] Wenn zum Beispiel ein Wohngeldantrag bearbeitet wird und ein Algorithmus darüber entscheidet, ob der Antrag akzeptiert oder abgelehnt wird, gibt es für die Antragstellenden keine Möglichkeit, das Ergebnis zu überprüfen oder Rückfragen zu stellen. Um eine Überprüfung oder Rückfragen zu ermöglichen, müsste entweder einer der Verwaltungsmitarbeitenden den Antrag händisch bearbeiten (dann könnte man sich allerdings den Einsatz des Algorithmus sparen), oder das algorithmische Ergebnis müsste einen Hinweis darauf enthalten, warum ein Wohngeldantrag abgelehnt wird.

Ein weiterer Unterschied zwischen menschlichen und algorithmischen Empfehlungen ist, dass das algorithmische Ergebnis objektiver sein kann als eine menschliche Empfehlung, wenn es gelingt, den Algorithmus mit verzerrungsfreien Daten zu trainieren. Zum Beispiel für eine Berufsempfehlung oder die Auswahl von neuem Personal sollten Identifikatoren wie Geschlecht und Herkunft irrelevant sein. Ein Mensch kann solche persönlichen Indikatoren seines Gegenübers nicht ignorieren und kann auch seine eigenen unbewussten Vorurteile nicht leicht ignorieren; ein Algorithmus hingegen kann so trainiert werden, dass Vorurteile gar keine Rolle spielen. Das würde jedoch eine sorgfältige Selektion der zum Training verwendeten Datensätzen erfordern und ist in manchen Anwendungsfällen gar nicht möglich.[801] Im Falle einer Berufsempfehlung wäre es möglich Identifikatoren wie Geschlecht oder Herkunft einer Person aus den Datensätzen zu löschen und somit bei dem Erstellen der Empfehlung zu ignorieren. Entscheidende Faktoren könnten dann unter anderem das persönliche Interesse und Leistungsdaten wie z.B. Schulnoten sein – allerdings sind Schulnoten in den meisten Fällen auch schon nicht frei von Verzerrungen.

Wenn also Datennutzende mit algorithmischen Ergebnissen konfrontiert werden, besteht die Gefahr, dass diese verzerrt sind und gleichzeitig keine Rückfragen gestellt werden können. Die algorithmischen Ergebnisse haben aber in den meisten Anwendungsfällen einen großen Einfluss auf die Datennutzenden, da diese womöglich

800 Buijsman, Veluwenkamp 2023:545; van den Hoven 1998:104f.
801 Mehr dazu unter 7.2.3

sogar epistemisch abhängig sind von den Ergebnissen. Deshalb stellt sich an dieser Stelle die Frage, wie es vermeidbar ist, dass Datennutzende möglicherweise verzerrten Ergebnissen ausgesetzt werden, denen sie Glauben schenken müssen, weil die Ergebnisse nicht überprüfbar sind. Denn so können womöglich falsche Überzeugungen entstehen. Es gibt in der Epistemologie verschiedene Ansprüche an Überzeugungen: Sie sollen wahr sein[802], oder sie sollen gerechtfertigt sein[803], oder sie sollen sowohl wahr als auch gerechtfertigt sein und somit als Wissen gelten können.[804] Doch an dieser Stelle ist es gar nicht notwendig, sich in diese sehr umfangreiche Debatte zu vertiefen. Denn für diese Arbeit ist lediglich relevant, dass Datennutzende nicht mit algorithmischen Ergebnissen konfrontiert werden sollen, die verzerrt (biased) sind. Dafür gibt es zwei Optionen: Entweder die Algorithmen sind verlässlich, oder die Ergebnisse können auf mögliche Verzerrungen überprüft werden.

7.1.1 Verlässlichkeit

Es gibt verschiedene Ansätze, um zu entscheiden, wann eine Überzeugung gerechtfertigt ist. Einer dieser Ansätze ist der Prozess-Reliabilismus. Einer der Hauptvertreter des Prozess-Reliabilismus, der Epistemologe Alvin Goldman, formuliert den Ansatz folgendermaßen: „The justificational status of a belief is a function of the reliability of the process or processes that cause it, where (…) reliability consists in the tendency of a process to produce beliefs that are true rather than false."[805] Das heißt, eine Überzeugung ist dann gerechtfertigt, wenn sie durch einen zuverlässigen Prozess erzeugt wird. Ein solcher Prozess ist zuverlässig, wenn er dazu neigt, wahre Überzeugungen zu produzieren.[806]

Das Konzept des Prozess-Reliabilismus bleibt vage in Bezug darauf, wie verlässlich ein Prozess der Überzeugungsbildung sein

802 Weiner 2005
803 Littlejohn 2009
804 Hawthorne, Stanley 2008:578; Ichikawa, Steup 2018; Buijsman, Veluwenkamp 2023:547
805 Goldman 1979:10
806 Ebd.:10, 14, 18

muss, um gerechtfertigte Überzeugungen hervorzubringen: „It does seem clear, however, that *perfect* reliability isn't required. Belief-forming processes that *sometimes* produce error still confer justification. It follows that there can be justified beliefs that are false."[807] Nach diesem Ansatz ist es also möglich, falsche Überzeugungen zu haben und trotzdem ist es gerechtfertigt, dass man daran glaubt. Goldman bleibt auch vage in Bezug auf die Frage, ob sich die Zuverlässigkeit eines Prozesses nur auf „(...) das Verhältnis von wahren zu falschen Überzeugungen in allen bisherigen und zukünftigen Anwendungen dieses Prozesses (...)"[808] bezieht, oder ob auch die Anwendungen in Situationen relevant sind, die hätten eintreten können.[809] Der Metaphilosoph Steffen Koch plädiert für einen Kompromiss, bei dem „(...) aktuale Anwendungen sowie Anwendungen in bestimmten nahen möglichen Welten Beachtung finden."[810] So stützt sich die Einschätzung der Verlässlichkeit eines Prozesses der Überzeugungsbildung nicht nur auf zu beobachtende Ergebnisse, sondern auch auf mögliche Ergebnisse beim Einsatz des Prozesses zu anderen Zwecken.

Der Prozess-Reliabilismus wird oft von einem Externalismus begleitet. Externalist:innen sind sich darüber im Klaren, dass die Erfahrungen eines Individuums teilweise durch seine Umwelt bestimmt werden. Diese Erfahrungen wiederum sind Glaubensquellen und beeinflussen daher den Glaubensbildungsprozess. Die Rechtfertigung einer Überzeugung hängt hier von der Zuverlässigkeit der Quellen der Überzeugung ab. Deshalb ist es wichtig zu fragen, ob Quellen wie Erinnerungen, Wahrnehmungen und introspektive Zustände zuverlässig sind.[811]

Laut Goldman sehen wir Überzeugungen als gerechtfertigt an, wenn sie durch Prozesse gebildet werden, die wir für zuverlässig halten. Unsere Annahmen darüber, welche glaubensbildenden Prozesse zuverlässig sind, können fehlerhaft sein, aber das habe keinen Einfluss auf deren Angemessenheit.[812] Um eine Überzeugung als

807 Goldman 1979:11
808 Koch 2019:170
809 Goldman 1979:11
810 Koch 2019:171
811 Comesaña 2010:571; Koch 2019:170; Steup, Neta 2020
812 Goldman 1979:18

Wissen bezeichnen zu können, reiche es jedoch nicht, von falschen Überzeugungen mit Hilfe verlässlicher Begründungsverfahren zu einer Schlussfolgerung zu kommen. Damit von Wissen gesprochen werden kann, müsse auch die gesamte Geschichte des Prozesses verlässlich sein und damit also auch die Überzeugungen, von denen man ausgeht.[813] Laut Goldman ist aber eine Überzeugung auch dann noch gerechtfertigt, wenn man sich an die ursprüngliche Begründung, die einen auf verlässliche Weise zu dieser Überzeugung gebracht hat, nicht mehr erinnern kann.[814]

Interessanterweise beschreibt Goldman einen solchen Prozess der Überzeugungsbildung (belief-forming process) so, wie wir heute den Prozess eines Algorithmus beschreiben würden: „Let us mean by a ‚process' a *functional operation* or procedure, i.e., something that generates *mapping* from certain states – 'inputs' – into other states – 'outputs'."[815] Mit „outputs" meint Goldman nicht etwa algorithmische Ergebnisse, sondern Zustände, in denen man zu einem bestimmten Zeitpunkt an eine Aussage glaubt. Eine solche gerechtfertigte Überzeugung resultiert aus Operationen der kognitiven Fähigkeiten, das heißt: „(...) 'information-processing' equipment *internal* to the organism."[816] Gemeint ist hier der Mensch als Organismus, der im Prozess der Überzeugungsbildung Informationen kognitiv verarbeitet. Aber genauso gut könnte sich dieser Begriff der Informationsverarbeitung heute auf den Vorgang eines Algorithmus beziehen.

Ebenso lässt sich der folgende Satz vom Menschen auf Maschinen übertragen: „A reasoning procedure cannot be expected to produce true belief if it is applied to false premises."[817] Wenn Menschen von falschen Annahmen ausgehen, können sie nicht zu wahren Überzeugungen gelangen – auch nicht mit Hilfe eines verlässlichen Prozesses. Gleichermaßen kann auch ein Algorithmus nicht auf Basis von verzerrtem Input einen nicht-verzerrten Output erzeugen – selbst wenn der Prozess der Informationsverarbeitung verlässlich ist. Goldman nennt einen Prozess bedingt verlässlich, wenn ein ausreichen-

813 Goldman 1979:15f.
814 Ebd.:15
815 Ebd.:11
816 Ebd.:13
817 Ebd.

der Anteil seiner Output-Überzeugungen wahr ist, vorausgesetzt, seine Input-Überzeugungen sind wahr.[818]

Die Technologiephilosophen Juan Manuel Durán und Nico Formanek haben, angelehnt an den Prozess-Reliabilismus, einen computergestützten Reliabilismus (computational reliabilism) entwickelt. Dieser besagt, es sei gerechtfertigt, dass Forscher von den Ergebnissen ihrer Simulationen überzeugt sind, wenn es einen zuverlässigen Prozess (d.h. die Computersimulation) gibt, der in den meisten Fällen wahre Ergebnisse hervorbringt.[819] Die Autoren machen die Zuverlässigkeit von Computersimulationen an Verifikations- und Validierungsmethoden, Robustheitsanalysen für Computersimulationen, Geschichten (un)erfolgreicher Implementierungen sowie Expert:innenwissen fest.[820] Die Zuverlässigkeit wird also durch formale Methoden und Leistungsvergleiche sichergestellt. Für erfolgreiche Implementierungen ist es außerdem wichtig, dass der Algorithmus in Übereinstimmung mit den epistemischen Standards eines Bereichs verwendet wird.[821] Während es Durán und Formanek um die Verlässlichkeit von Computersimulationen geht, übertragen Durán und Jongsma die Idee der Verlässlichkeit auf den Einsatz von KI in der Medizin. Von den formalen Methoden für die Feststellung der Zuverlässigkeit übernehmen Letztere nur das Expert:innenwissen und fügen die Transparenz als Methode hinzu, verstanden als ein Prozess, der das Innenleben von Black-Box-Algorithmen offenlegt.[822]

Gegen die neue Methode von Durán und Jongsma würde die Kantianerin Maya Krishnan vermutlich einwenden, dass durch die Transparenz der Vorteil des Reliabilismus verloren geht. Denn der Reliabilismus liefert eine plausible Erklärung dafür, warum man sich auf die Ergebnisse der (kognitiven) Informationsverarbeitung verlassen kann, auch wenn es keine detaillierte wissenschaftliche Theorie gibt: weil das System zuverlässig das richtige Ergebnis produziert: „In some cases, a record of success can stand in for a precise account

818 Goldman 1979:13
819 Durán, Formanek 2018:654
820 Ebd.:656ff.
821 Grote et al. 2024:6
822 Durán, Jongsma 2021:332

of the inner workings of an instrument."[823] Es kann also gerade dank des Reliabilismus eine genaue Beschreibung der inneren Funktionsweise eines Instruments ersetzt werden durch seine Erfolgsbilanz.

Laut Grote et al. kann die Zuverlässigkeit von Algorithmen bewertet werden, indem ihre Robustheit getestet, ihre Vorhersagegenauigkeit überprüft und ihre internen Entscheidungsprozesse bewertet werden. Ein Algorithmus gilt dabei als robust, wenn seine Leistung relativ stabil ist. Seine Vorhersagegenauigkeit kann anhand seiner Ergebnisse getestet werden. Doch das reicht laut Grote et al. noch nicht aus, da es möglich ist, dass ein Algorithmus eine hohe Vorhersagegenauigkeit erreicht, dies aber auf der Grundlage von Merkmalen, die wir Menschen für falsch halten. So gibt zum Beispiel ein von den bereits erwähnten Informatikern Ribeiro et al. programmierter Algorithmus aus, dass auf einem Bild ein Wolf zu sehen ist, wenn Schnee (oder ein heller Hintergrund im unteren Bildbereich) vorhanden ist, und wenn auf dem Bild kein Schnee zu sehen ist, gibt der Algorithmus an, es sei ein Husky zu sehen, unabhängig von der Farbe, der Position oder der Haltung des Tieres. Dies liegt daran, dass der Algorithmus ausschließlich mit Bildern von Wölfen im Schnee und solchen von Huskys ohne Schnee trainiert wurde.[824] Grote et al. ziehen daraus folgenden Schluss: „In the case of the wolf v. Husky classifier, the model is unreliable, because it relies on the wrong features (…)."[825] Solange auf allen neuen Bildern von Wölfen Schnee im Hintergrund ist und auf allen Bildern von Huskys kein Schnee, wäre die Vorhersagegenauigkeit des Algorithmus sehr hoch – allerdings auf Grund von Merkmalen, die nichts mit den jeweiligen Tieren zu tun haben. Diese unerwünschten Korrelationen, die der Algorithmus während des Trainings erkannt hat, sind sehr schwer zu identifizieren, wenn man sich nur die Rohdaten und Vorhersagen ansieht.[826] Deshalb ist laut Grote et al. die Bewertung des internen Entscheidungsprozesses von Algorithmen ein zentraler Ansatz zur Beurteilung ihrer Zuverlässigkeit.[827]

823 Krishnan 2020:496
824 Ribeiro et al. 2016:1142
825 Grote et al. 2024:6
826 Ribeiro et al. 2016:1142
827 Grote et al. 2024:6

Die Erfolgsbilanz von Algorithmen ist für Datennutzende, die mit einem algorithmischen Ergebnis konfrontiert werden, nicht leicht zu überprüfen, deshalb verlassen sie sich diesbezüglich auf die Entwickler:innen des Systems. Der Reliabilismus geht in Bezug auf algorithmische Ergebnisse davon aus, dass es gerechtfertigt wäre, wenn die Datennutzenden die Ergebnisse des Systems in ihre praktischen Überlegungen einbeziehen – angenommen den Datennutzenden wurde von den Entwickler:innen des Systems korrekterweise mitgeteilt, dass das System insgesamt zuverlässig ist.[828] Die Epistemologin Katherine Dormandy wendet ein: „Allerdings ist anzumerken, dass es ohne Belege alles andere als klar ist, ob eine gewisse Überzeugung überhaupt zuverlässig gebildet wurde."[829] Die Entwickler:innen brauchen mindestens Informationen darüber, ob das System in verschiedenen Anwendungsfällen mehr wahre als falsche Ergebnisse erzeugt hat, um das System als zuverlässig einstufen zu können.

Die Fokussierung auf die Erfolgsbilanz eines Prozesses bedeutet zudem, dass es bei einem Algorithmus, der als zuverlässig gilt, gerechtfertigt wäre, von seinen Ergebnissen überzeugt zu sein, selbst wenn sie falsch sind. Das Gleiche gilt andersherum: Gilt ein Algorithmus als unzuverlässig, wäre es nicht gerechtfertigt, wenn Datennutzende von den algorithmischen Ergebnissen überzeugt wären.[830] Wenn man also herausfinden möchte, ob man algorithmischen Ergebnissen gerechtfertigter Weise glauben kann, ist man angewiesen auf das Urteil der Programmierenden und Systembetreibenden über die Verlässlichkeit des Algorithmus. Die Programmierenden und Systembetreibenden wissen bei hochkomplexen Algorithmen aber auch nicht genau, wie das System funktioniert und machen dementsprechend seine Zuverlässigkeit an seiner Erfolgsbilanz fest.

Es fragt sich nun, ob es ein Problem ist, von algorithmischen Ergebnissen überzeugt zu sein, wenn sie falsch sind. Wenn im Datenökosystem Transparenz fehlt, können Datennutzende durch algorithmische Ergebnisse zu falschen Überzeugungen gelangen. Gelangen Datennutzende auf diesem Wege zu falschen Überzeugungen, können sie Schwierigkeiten haben, epistemisch wertvolle Zustände

828 Buijsman, Veluwenkamp 2023:548
829 Dormandy 2019:184
830 Buijsman, Veluwenkamp 2023:548

zu erreichen.[831] Epistemisch wertvoll ist das, was in einer relevanten Beziehung zum Ziel der Wahrheit steht. Dabei ist es das epistemische Ziel, genaue und umfassende Überzeugungen zu haben.[832] Wahre Überzeugungen sind laut dem Epistemologen William P. Alston notwendig zum Überleben und zum Vermeiden von Chaos: „(…) it is important for human flourishing to be guided by correct rather than incorrect suppositions about how things are, where this is of interest or importance to us (…)"[833] Da es für das menschliche Wohlergehen wichtig ist, sich von richtigen und nicht von falschen Überzeugungen leiten zu lassen, sollte das Erzeugen von falschen Überzeugungen durch algorithmische Ergebnisse möglichst vermieden werden. Deshalb ist es ein Problem des Reliabilismus, dass es gerechtfertigt sein kann, von algorithmischen Ergebnissen überzeugt zu sein, selbst wenn sie falsch sind.

Ein weiteres Problem bei diesem Ansatz ist, dass der Reliabilismus äußeren Faktoren und Gründen gegenüber blind ist. Denn der Prozess der Informationsverarbeitung ist für externe Faktoren empfindlich, aber er bringt Überzeugungen hervor, ohne dass das Subjekt diese Empfindlichkeit bewusst reflektiert.[834] Wenn also Überzeugungsquellen zuverlässig waren und der Prozess der Überzeugungsbildung wahre Überzeugungen hervorgebracht hat, wird erwartet, dass dieser zuverlässig ist und automatisch immer wahre Überzeugungen hervorbringt. Dagegen spricht laut dem Technologiephilosophen Stefan Buijsman vor allem eins: „(…) algorithms will perform worse in outlier cases, even when they are generally accurate."[835] So mag ein Algorithmus als verlässlich gelten, weil er oft genug richtig lag. Aber es bleibt trotzdem unklar, ob der Algorithmus in allen Fällen gute Entscheidungen trifft. Um Einzelfälle zu erkennen, in denen der Algorithmus verzerrte Ergebnisse bereitstellt, müssten sehr viele Tests für alle möglichen Ausnahmen geprüft werden. Der Algorithmus von Amazon galt als verlässlich, bis deutlich wurde, dass Bewerberinnen systematisch aussortiert wurden. Buijsman hat einen Lösungsvorschlag für das Dilemma: „Understanding why the algo-

831 Harris 2022:137
832 Foley 1993:102, 198
833 Alston 2005:30
834 Comesaña 2010:571
835 Buijsman 2023:203

rithm presents a certain output may help to gain a more fine-grained justification, and avoid mistakes that would arise if the algorithm is given blanket trust."[836] Deshalb soll im Folgenden der Ansatz des Evidentialismus betrachtet werden.

7.1.2 Überprüfbarkeit

Ein anderer Ansatz um zu entscheiden, wann eine Überzeugung gerechtfertigt ist, ist der Evidentialismus. Zwei der Hauptvertreter der Position, die Epistemologen Earl Conee und Richard Feldman, formulieren diese so: „What we call evidentialism is the view that the epistemic justification of a belief is determined by the quality of the believer's evidence for the belief."[837] Die epistemische Rechtfertigung einer Überzeugung hängt bei diesem Ansatz nicht von den kognitiven Fähigkeiten der Menschen oder von den kognitiven Prozessen oder Praktiken der Informationsbeschaffung ab, die zu der Überzeugung geführt haben – sie hängt allein von den Beweisen ab, die man für sie hat. Eine Überzeugung ist laut Conee und Feldman dann „gut begründet" (well-founded), wenn sie sowohl epistemisch gut gestützt als auch auf die richtige Art und Weise zustande gekommen ist. Ob eine Überzeugung „gut begründet" ist, hängt ab von der Evidenz, die man hat, und der Evidenz, die man bei der Bildung der Überzeugung verwendet. Man muss also „(...) *aufgrund der entsprechenden Belege* (...)"[838] zu einer Überzeugung kommen, damit die Überzeugung gemäß dem Evidentialismus gerechtfertigt ist.[839]

Der Evidentialismus bietet auch eine Erklärung dafür, woher wir wissen, wann wir zu wahren Überzeugungen gelangt sind. Es reicht nicht, an die Wahrheit der Überzeugung zu glauben, um ihre Wahrheit festzustellen.[840] Um herauszufinden, ob eine Überzeugung wahr ist, braucht man Beweise für diese Überzeugung.[841] Als Beweis reicht es jedoch nicht, eine Rechtfertigung für den Glauben an die

836 Buijsman 2023:203
837 Conee, Feldman 2004:83
838 Dormandy 2019:179
839 Conee, Feldman 2004:93
840 Ebd.:251
841 David 2001:154, 160, 163; Conee, Feldman 2004:249

Wahrheit der Überzeugung vorzubringen.[842] Denn man könnte gerechtfertigter Weise den Aussagen von Ärzt:innen Glauben schenken und dadurch dennoch zu falschen Überzeugungen gelangen – weil sich die Ärzt:innen zum Beispiel irren. Laut Conee und Feldman brauchen wir stattdessen eine Rechtfertigung für die Überzeugung selbst: „A proposition is epistemically justified to someone when it is evident to the person that the proposition is true."[843] Was epistemisch gerechtfertigt werden muss, um zu wissen, ist nicht „die Überzeugung" im Sinne von: der Glaube an die Aussage. Es ist „die Überzeugung" im Sinne von: die geglaubte Aussage.[844] Um herauszufinden, ob eine Überzeugung im Sinne einer geglaubten Aussage wahr ist, brauchen wir also Beweise für die Wahrheit dieser Überzeugung.

Solche Beweise können in Form von Propositionen angeführt werden. Eine Proposition kann mit Hilfe der Vernunft oder aus dem Zeugnis von anderen abgeleitet werden.[845] Da wir nicht alles selbst erfahren können und uns nicht alles nur aus unserer Vernunft ableiten können, müssen wir uns manchmal auf das Zeugnis anderer verlassen. Allerdings reicht das Zeugnis anderer nur dann als Beweis für die Wahrheit einer Überzeugung, wenn man diese nicht selbst vernünftig herleiten kann oder die notwendigen Erfahrungen nicht selbst machen kann.[846] Außerdem ist das Zeugnis anderer nur dann eine Beweisquelle, wenn die anderen eine zuverlässige Quelle darstellen.[847] Wenn man zum Beispiel mit chronischen Bauchschmerzen zu einer Ärztin geht, macht man die Erfahrung selbst, kann sich aber die Ursache als Laie nicht vernünftig herleiten – deshalb muss man sich für eine Erklärung der chronischen Bauchschmerzen auf die Ärztin verlassen. Damit ihr Zeugnis als Beweisquelle für eine wahre Überzeugung dienen kann, muss sie als Quelle zuverlässig sein. In diesem Beispiel bedeutet das, dass die Ärztin durch ein Medizinstudium ausgebildet worden sein muss, ihr die Approbation erteilt wurde und sie sich an den Ärzte-Kodex hält.

842 David 2001:160; Conee, Feldman 2004:251
843 Conee, Feldman 2004:252f.
844 Ebd.:252
845 Ebd.:238; Steup, Neta 2020
846 Steup, Neta 2020
847 Conee, Feldman 2004:225; Steup, Neta 2020

Wenn allerdings Beweise nur solche für wahr oder wahrscheinlich gehaltene Propositionen sein können, entsteht das Problem des infiniten Regresses, weil laut Dormandy „(...) Belege selber durch weitere Belege gestützt werden müssen."[848] Eine Möglichkeit zur Auflösung dieses Problem ist Conee und Feldman zufolge das Akzeptieren von repräsentationalen Erfahrungen, wie zum Beispiel Wahrnehmungen und Erinnerungen, als Beweise: „(...) I do not wish to rule out the possibility that experiences or perceptual states can count as evidence as well."[849] Solche Erfahrungen sind selbst nicht begründbar, aber leisten trotzdem einen wichtigen Beitrag zum Sammeln von Beweisen und zum Bilden einer Überzeugung. Über Wahrnehmungen kann laut Dormandy die Annahme gemacht werden, „(...) dass ihnen relativ viel epistemisches Gewicht zugeschrieben werden kann, sofern es keine Gegenbelege (*defeaters*) gibt."[850] Belege stützen eine Überzeugung dann, wenn „(...) diese Überzeugung die beste verfügbare Erklärung für die Belege bietet."[851] Das heißt, man nimmt zum Beispiel wahr, dass in der Praxis der Ärztin eine Approbationsurkunde mit dem Namen der Ärztin an der Wand hängt. Dieser Beleg stützt die Überzeugung, dass die Ärztin ihr Medizinstudium erfolgreich abgeschlossen hat. Die Approbationsurkunde könnte auch gefälscht sein, aber die plausibelste Erklärung für diese Wahrnehmung (Beleg) ist, dass die Ärztin ihr Medizinstudium abgeschlossen hat und dann einen erfolgreichen Antrag auf Approbation gestellt hat.

Da repräsentationale Erfahrungen auch als Belege gelten, geht der Evidentialismus mit einem Internalismus einher. Es gibt einerseits den mentalistischen Internalismus (mentalism) und andererseits den Zugangsinternalismus (accessibilism). Denn die Dinge, die eine Überzeugung rechtfertigen, sind entweder mentale Zustände (mentalistischer Internalismus) oder sie sind bei Reflexion erkennbar (Zugangsinternalismus). Den Zugangsinternalismus definieren Conee und Feldman folgendermaßen: „What we shall call "accessibilism" holds that the epistemic justification of a person's belief is determined by things to which the person has some special sort

848 Dormandy 2019:180
849 Conee, Feldman 2004:225f.
850 Dormandy 2019:183
851 Ebd.:181

of access."[852] Gegen diese zweite, logisch stärkere Version des Internalismus spricht jedoch, dass er den Menschen zu viel abverlangt. Conee und Feldman plädieren auf Grund der Einfachheit und Klarheit für den mentalistischen Internalismus: „(...) [mentalist] internalism is the view that a person's beliefs are justified only by things that are internal to the person's mental life."[853] Die Überzeugungen einer Person sind demnach durch Beweise gerechtfertigt, welche die Person erlebt oder reflektiert.[854]

Das Problem beim mentalistischen Internalismus ist, dass unklar ist, wann die erlebte oder reflektierte Evidenz geeignet ist, eine Überzeugung zu rechtfertigen. Ohne Regeln für die Qualität der Evidenz könnte jede Erinnerung oder Intuition als Beweis für die Wahrheit eines Glaubens gelten. Dies macht die Beweise und den gesamten Rechtfertigungsprozess sehr subjektiv und schwer überprüfbar.[855] Wenn auf der Grundlage von algorithmischen Ergebnissen Überzeugungen gebildet werden, dann werden diese Ergebnisse dabei als Belege für diese Überzeugung betrachtet. Dieser Umstand verdeutlicht das Problem, dass es so kein Qualitätskriterium für Belege gibt. Denn algorithmische Ergebnisse können verzerrt sein, ohne dass Datennutzende und sogar auch Datenverarbeitende sich dessen bewusst sind. In dem Fall kommen Datennutzende „aufgrund der entsprechenden Belege"[856] zu einer Überzeugung, sodass die Überzeugung gemäß dem Evidentialismus gerechtfertigt ist.[857] Doch auch bei falschen Belegen in Form von verzerrten algorithmischen Ergebnissen wäre die Überzeugung gerechtfertigt, da es scheint, als hätte man Belege für die Wahrheit der Überzeugung, und doch ist sie womöglich nicht wahr.

852 Conee, Feldman 2004:55
853 Ebd.:55
854 Dormandy 2019:179; Steup, Neta 2020
855 Comesaña 2010:571
856 Dormandy 2019:179
857 Conee, Feldman 2004:93

7.1.3 Ein Kompromiss

Der Epistemologe Juan Comesaña fasst zusammen, dass der Reliabilismus nicht ohne Beweise kann und der Evidentialismus nicht ohne Verlässlichkeit: „Reliability without evidence is blind, evidence without reliability is empty."[858] Um eine Lösung für die unterschiedliche Qualität der Evidenz im Evidentialismus und den Mangel an Reflexion im Reliabilismus zu finden, hat Comesaña beide Ansätze miteinander kombiniert. Sein sogenannter evidentialistischer Reliabilismus kann wie folgt definiert werden: „(...) to be justified (…) is to believe in accordance with one's evidence, and one's beliefs accord with one's evidence if and only if that evidence is reliably connected to the truth of those beliefs."[859] Eine Überzeugung ist also gerechtfertigt, wenn sie auf der Grundlage von Beweisen gebildet wird. Die Evidenz wiederum ist geeignet, einen Glauben zu rechtfertigen, wenn sie zuverlässig ist.[860] Auf diese Weise wird das Problem des Evidentialismus gelöst, da es jetzt eine Bedingung für die Qualität der Evidenz gibt. Gleichzeitig wird das Problem des Reliabilismus gelöst, weil wir uns nicht auf einen automatischen Glaubensbildungsprozess verlassen müssen, sondern bewusst über die Evidenz nachdenken, die wir erfahren. Indem wir über die Beweise und ihre Zuverlässigkeit nachdenken, können wir die Überzeugungen, die wir bilden, rechtfertigen.

So muss man sich nicht darauf verlassen, dass der Algorithmus meistens wahre Ergebnisse erzeugt, sondern man sammelt Belege dafür. Gleichzeitig muss man sich nicht fragen, welche der Belege qualitativ hochwertig genug sind, um eine Überzeugung rechtfertigen zu können, da das Kriterium die Verlässlichkeit ist. Der evidentialistische Reliabilismus löst allerdings in der Praxis auch nicht alle Probleme. Denn wenn wir nun also mit Hilfe eines Algorithmus zu einer Überzeugung gelangen, brauchen wir Beweise für die Wahrheit dieser Überzeugung, und diese Beweise müssen verlässlich sein. Es ist aber auf Grund der Komplexität des Algorithmus oft schwierig, Beweise außerhalb der algorithmischen Ergebnisse selbst zu finden oder sie mittels unserer Vernunft abzuleiten. Die Tech-

858 Comesaña 2010:571
859 Ebd.:595
860 Ebd.:584

nologiephilosophen Stefan Buijsman und Juan M. Durán schreiben über Machine Learning Models: „(…) their epistemic opacity makes it difficult to extract causal relations and even to determine the reliability and robustness of their predictions."[861] Der algorithmische Prozess ist somit im Wesentlichen epistemisch undurchsichtig, weil es Menschen auf Grund ihrer Natur unmöglich ist, alle epistemisch relevanten Elemente des Prozesses zu kennen.[862] Deshalb müssen wir uns laut dem Epistemologen und Bioethiker John Hardwig in solchen Fällen auf das Zeugnis anderer verlassen: „(...) one can have good reasons for believing a proposition if one has good reasons to believe that *others* have good reasons to believe it (…)."[863] Laut Grote et al. generieren die Algorithmen selbst dieses Zeugnis in Form von ihren Ergebnissen[864], aber oft ist nicht klar, ob dieses Zeugnis auf „guten Gründen" beruht. Über die Gründe der Algorithmen müsste dementsprechend eine Person Zeugnis ablegen, die deren Vorgehen verstehen und nachvollziehen kann, wie die Ergebnisse zustande gekommen sind. Ein Algorithmus selbst gibt nicht von sich aus an, auf welchem Weg sein Output entstanden ist, und sowohl die Programmierenden als auch die Betreibenden des Algorithmus wissen das bei erhöhter Komplexität auch nicht genau.

Bei einem einfachen Wenn-Dann-Algorithmus wissen die Programmierenden und die Betreibenden, auf Grund welcher Entscheidungskriterien der Algorithmus zu seinem Ergebnis kommt, weil sie die Kriterien vorgegeben haben. Sie können also verlässliches Zeugnis darüber ablegen, dass der Algorithmus den Wohnantrag von Person A abgelehnt hat, weil Person A nicht alle gesetzlich notwendigen Bedingungen erfüllt. Programmierende und Betreibende wissen in dem Fall, dass weder Geschlecht noch Religion von Person A ausschlaggebend für die Entscheidung gewesen sind. Sie haben dem Algorithmus die Kriterien, zum Beispiel Wohnungsgröße und Einkommen, vorgegeben und können demnach Zeugnis darüber ablegen, wie die Ergebnisse des Algorithmus zustande kommen. Mit Hilfe dieses verlässlichen Zeugnisses können Datennutzende Beweise für die Wahrheit ihrer durch den Algorithmus verursachten Überzeugungen sammeln.

861 Buijsman, Durán 2024:466
862 Humphreys 2009:618
863 Hardwig 1985:336
864 Grote et al. 2024:5

Handelt es sich aber um ein komplexes KI-System, bei dem selbst für die Programmierenden nicht mehr nachvollziehbar ist, wie der Algorithmus zu seinen Ergebnissen kommt, gibt es folglich niemanden, der verlässlich Zeugnis darüber ablegen kann. Verlässlich sind die Algorithmen nur, wenn jemand nachvollziehen kann, wie die Ergebnisse zustande kommen – das ist nur bei Wenn-Dann-Algorithmen einfach möglich. Die Überprüfung der Ergebnisse auf mögliche Verzerrungen, als alternative Absicherung der durch sie gebildeten Überzeugungen, ist ebenfalls deutlich leichter, wenn das Vorgehen des Algorithmus bekannt ist. Deshalb ist es an dieser Stelle wichtig, sich mit der Erklärbarkeit und Interpretierbarkeit von Algorithmen auseinanderzusetzen. Denn dem Informationswissenschaftler Michael Winikoff zufolge ist eine der wesentlichen Voraussetzungen für angemessenes Vertrauen in ein autonomes System die Fähigkeit zu erklären, warum das System eine Entscheidung getroffen hat.[865]

7.2 Erklärbarkeit und Interpretierbarkeit

Um Beweise zu sammeln für die Verlässlichkeit von algorithmischen Ergebnissen können diese erklärt oder interpretiert werden. Sowohl Erklärung als auch Interpretierbarkeit sind erkenntnistheoretische Begriffe. Wie unter 2.1 erörtert, sind sie jedoch nicht gleichzusetzen. Ein Algorithmus ist interpretierbar, wenn seine Ergebnisse bei der Inspektion des Algorithmus verstanden werden können. Ein Algorithmus ist erklärbar, wenn die Ergebnisse durch einen separaten Prozess verstanden werden können, indem ein anderer Algorithmus zusammenfasst, wie der zu erklärende Algorithmus vorgeht.[866] Interpretierbarkeit ist vor allem bei Wenn-Dann-Algorithmen möglich, da hier zumindest die Entwickler:innen nachvollziehen können, wie die Ergebnisse zustande gekommen sind. Deshalb empfiehlt die Informatikern Cynthia Rudin hingegen, statt Black-Box Algorithmen interpretierbare Algorithmen zu verwenden.[867] Wenn aber doch komplexere Black-Box Algorithmen verwendet wurden, werden Erklärungen gebraucht.

865 Winikoff 2018:8
866 Russel, Norvig 2021:729
867 Rudin 2019; Buijsman 2023:206

Buijsman zufolge gibt es keine allgemein anerkannte Definition von Erklärung (explanation) und auch keine Möglichkeit, die Qualität verschiedener Erklärungen zu vergleichen.[868] Aber er meint, wenn wir Erklärungen fordern, fragen wir „Warum?" – also zum Beispiel „Warum gibt der Algorithmus diese Ergebnisse aus?". Buijsman wird noch konkreter und behauptet, dass Erklärungen Was-wäre-wenn-es-anders-gewesen-wäre-Fragen sind. Er nennt sie kontrastive Warum-Fragen (contrastive why-questions) und meint damit, dass den Datennutzenden ein alternatives Ergebnis präsentiert werden sollte und eine Eingabe, die zu diesem Ergebnis führen würde – nur so ließe sich die Funktionsweise eines Algorithmus vollständig erklären. Je mehr solche kontrastiven Warum-Fragen beantwortet würden, desto besser sei die Erklärung.[869]

Neben diesem interventionistischen Ansatz, der sich darauf konzentriert zu klären, wie das Ergebnis aussehen würde, wenn der Input ein anderer wäre, gibt es noch zwei weitere Ansätze zur Definition von Erklärung: den vereinheitlichenden und den mechanistischen Ansatz. Beim vereinheitlichenden Ansatz geht es darum, möglichst allgemeine Argumentationsmuster zu identifizieren, die es ermöglichen, eine breite Palette von Ergebnissen abzuleiten. Im Idealfall schafft es eine Theorie, ein Regelwerk aufzustellen, welches ein breites Spektrum von Phänomenen zusammenfasst. Die Newtonschen Gesetze gelten als Ideal der Vereinheitlichung, weil sie eine einheitliche Erklärung von Bewegungen auf der Erde und im Weltall ermöglichen.[870] Mechanistische Ansätze hingegen beschreiben das Innenleben eines Systems mit so vielen kausal relevanten Details wie möglich. Die Umsetzung dieses Ansatzes ist bei komplexen KI-Systemen jedoch schwierig, da hier jeder der zahlreichen Parameter eine gewisse kausale Bedeutung für die Ausgabe haben könnte und somit ein vollständiges Modell der Funktionsweise sehr komplex wäre und deshalb nicht viel Aufschluss geben könnte.[871]

Eine Erklärung von algorithmischen Ergebnissen ist laut Goodman und Flaxman nur möglich, wenn der Algorithmus von einem Menschen in Worte gefasst und verstanden werden kann. Sie neh-

868 Buijsman 2022:564
869 Ebd.:573, 580f.; Buijsman 2023:206
870 Buijsman 2023:209
871 Ebd.:212f.

men an, dass dafür die Beziehungen zwischen den Eingabemerkmalen und den Ergebnissen dargestellt werden müssen, sodass deutlich wird, welche Merkmale die Ergebnisse in welcher Weise beeinflussen.[872] Krishnan fügt hinzu, dass durch eine Erklärung in der Lage sein sollte zu wissen, warum der Algorithmus genau diese Ausgabe erzeugt.[873] Sobald die Verhältnisse zwischen Input und Output auf diese Weise offengelegt werden, wären sowohl der vereinheitlichende Ansatz der Erklärung als auch der interventionistische Ansatz im Kontext von Algorithmen umsetzbar. Denn dann können einerseits möglichst allgemeine Argumentationsmuster identifiziert werden, die es ermöglichen, eine breite Palette von Ergebnissen abzuleiten (vereinheitlichender Ansatz).[874] Wenn zum Beispiel das Alter in Kombination mit dem Einkommen einen Einfluss auf die Ergebnisse hat, können mit dieser Regel sowohl Kreditverweigerungen als auch zugeschriebene Kreditwürdigkeit erklärt werden. Andererseits ist es möglich, mit Hilfe der Erkenntnisse über die Beziehung zwischen Input und Output Was-wäre-wenn-es-anders-gewesen-wäre-Fragen zu beantworten (interventionistischer Ansatz).[875] Denn wenn man weiß, welchen Einfluss welcher Input auf das Ergebnis hat, kann man ableiten, wie das Ergebnis aussehen würde, wenn der Input anders wäre.

Manche Algorithmen sind so komplex, dass es schwierig bis unmöglich ist, die Regeln zu verstehen, nach denen der Input mit dem Output verknüpft ist. Für solche Algorithmen haben Informatiker:innen Erklärungsmethoden entwickelt, um die Transparenz zu erhöhen. So gibt es zum Beispiel sogenannte kontrafaktische Erklär-Techniken (counterfactual XAI techniques), die Mittel zur Identifizierung von unterscheidenden Eigenschaften (difference-making properties) eines Algorithmus liefern. Laut Munch et al. kann das ausreichen, um das Verhalten des Algorithmus zu erklären.[876] Die Informatiker:innen Osbert Bastani, Carolyn Kim und Hamsa Bastani nennen ihre Erklärungsmethode „Model Extraction". Hier

872 Goodman, Flaxman 2017:55
873 Krishnan 2020:493
874 Buijsman 2023:209
875 Buijsman 2022:573, 580f.; Buijsman 2023:206
876 Munch et al. 2024:4; Für Hinweise auf Fallstricke der Modellinterpretation siehe Molnar et al. 2022.

wird eine einfachere Version des zu erklärenden Algorithmus mit expliziten Regeln für die Beziehung zwischen Input und Output erstellt. Daraufhin wird geprüft, ob der Output des einfacheren Algorithmus mit dem Output des undurchsichtigen Algorithmus übereinstimmt, wenn beide neuen Input erhalten. Auf diese Weise ist der komplizierte Algorithmus dann anhand von der einfacheren Version interpretierbar.[877]

Laut Buijsman können zum Beispiel Werkzeuge wie Feature Importance Maps oder Bias Detection aufdecken, wenn die algorithmischen Ergebnisse auf irrelevanten oder ungerechten Merkmalen beruhen: „Feature importance maps and counterfactuals can hint that the model looks at features that we think are irrelevant, even if they generally do not enlighten us as to why the model looks at those features."[878] Solche Verfahren und Methoden, die algorithmische Merkmale nachvollziehbar machen sollen, werden als erklärbare KI (explainable AI; XAI) bezeichnet. Andere Werkzeuge aus diesem Bereich können mit Hilfe von statistischer Analyse das Erkennen von Verzerrungen unterstützen und so auf die Gruppen hinweisen, die auf Grund von bestimmten Merkmalen benachteiligt und somit diskriminiert werden.[879] Daraufhin können solche Ungleichheiten manuell von menschlichen Expert:innen geprüft werden und Fehler können womöglich behoben werden. Deshalb können derartige erklärbare KI-Methoden auch dann nützlich sein, wenn sie keine eigentlichen Erklärungen liefern.[880]

Die Informatiker Anupam Datta, Shayak Sen und Yair Zick sind ebenfalls der Meinung, dass die Transparenz von Algorithmen verbessert werden kann, indem erklärt wird, warum eine bestimmte Empfehlung erzeugt wurde. Deshalb führen sie sogenannte „Quantitative Input Influence (QII) measures" ein, die den Grad des Einflusses von Inputs auf die Outputs von Systemen erfassen. Anhand von verschiedenen Variablen können Entscheidungen über Einzelpersonen und Gruppen erklärt werden, indem sie den gemeinsamen Einfluss einer Gruppe von Inputs (z.B. Alter und Einkommen) auf die

877 Bastani et al. 2017; Vredenburgh 2022:221
878 Buijsman 2023:212
879 Für einen Überblick über den Stand der Technik von XAI Methoden siehe Holzinger et al. 2022.
880 Buijsman 2023:212

Ergebnisse sowie den marginalen Einfluss einzelner Inputs innerhalb einer solchen Gruppe (z.B. Einkommen) quantifizieren.[881] Auch diese Methode von Datta et al. gilt als Werkzeug der erklärbaren KI (XAI-Tool). Hierdurch ist es auch bei undurchsichtigen Algorithmen möglich zu beobachten, wie sie bestimmte Inputs zuverlässig mit bestimmten Outputs in Verbindung bringen. Dadurch wissen wir zwar immer noch nicht, warum diese Korrelationen hergestellt werden. Trotzdem ist die Erklärung schon mal spezifischer, wenn nicht nur angegeben wird, welche Merkmale für ein algorithmisches Ergebnis wichtig sind, sondern es genaue Schätzungen darüber gibt, auf welche Weise diese Merkmale wichtig sind.

Der Digitalethiker Lauritz Aastrup Munch, der Technologiephilosoph Jens Christian Bjerring und der Digitalethiker Jakob Thrane Mainz veranschaulichen diesen Erklärungsprozess anhand des Beispiels eines durch ein algorithmisches System verweigerten Kredits. Wenn die Erklärungsmethoden (XAI-Tools) ergeben, dass das niedrige Einkommen einer Person als Input Einfluss hatte auf den Output des Algorithmus, dann wird impliziert, dass das Einkommen als Input zumindest eine Teilursache für die Entscheidung der Kreditverweigerung ist. Die Nennung des Einkommens als zentraler Faktor zur Erklärung einer algorithmischen Schätzung der Kreditwürdigkeit sagt etwas darüber aus, wie das Modell funktioniert. Da das Modell Daten aus der echten Welt reproduziert, spiegeln solche kausalen Beziehungen oft den Einfluss des Einkommens auf die Kreditwürdigkeit in der Vergangenheit wider. Laut Munch et al. besteht die Hoffnung, dass diese XAI-Tools zumindest teilweise ein kausales Verständnis von KI-Systemen ermöglichen und uns somit zumindest einige der Dinge liefern, die uns kausale Erklärungen normalerweise liefern.[882]

Die Technologiephilosophen Timo Freiesleben und Gunnar König, die sich auf Machine Learning spezialisiert haben, machen darauf aufmerksam, dass solche Erklär-Methoden jedoch nicht ohne Kontext hilfreich sind. Sie meinen, oft bleibe unklar, welchem Zweck diese Erklärungen dienen könnten, und unter welchen Bedingungen sie hilfreich sind. Ihrer Meinung nach sollte jede Erklärung mindestens einem präzisierten Zweck dienen: „A purpose is a goal humans

881 Datta et al. 2016:1f.
882 Munch et al. 2024:4f.

have in mind when they ask for explanations."[883] Je nach Ziel der Erklärung können unterschiedliche Methoden angemessen sein. Dabei sollen die Methoden nicht dazu genutzt werden, Datennutzende dazu zu bringen, den Algorithmen zu trauen, sondern sie sollen dazu beitragen, den Mechanismus der Algorithmen transparenter zu machen. Dabei stellen die Autoren in Frage, ob die Datennutzenden die Wirkweise von Algorithmen verstehen werden. Algorithmen sind komplex, Erklärungen hingegen sollen einfach sein und nicht jeden Aspekt eines Algorithmus wiedergeben. Eine mögliche Lösung dafür wäre es laut Freiesleben und König, verschiedene Arten von Erklärmethoden gleichzeitig einzusetzen, damit mehrere Aspekte des Algorithmus beleuchtet werden können.[884]

7.2.1 Erklärungen für algorithmische Ergebnisse

Es gibt jedoch einige Autor:innen, die in Frage stellen, ob wir überhaupt Erklärungen für algorithmische Ergebnisse benötigen. Buijsman hat vier dieser Argumentationen zusammengestellt und widerlegt. Als erstes widmet er sich dem Argument des KI-Ethikers Scott Robbins, dass nicht der Algorithmus erklärt werden müsse, sondern seine Entscheidungen. Wir bräuchten nur dann Erklärungen, wenn die Auswirkungen der Entscheidungen auf andere Menschen hoch sind oder sein können. So behauptet Robbins zum Beispiel, dass die Entscheidungen von Algorithmen, die Muttermale als gutartig oder bösartig einstufen, nicht erklärt werden müssen. Denn wenn ein Muttermal als bösartig eingestuft wird, sei die Folge dieser Entscheidung lediglich eine Biopsie. Ärzt:innen seien auch nicht verpflichtet, ihre Entscheidungen zu erklären, deshalb solle das auch beim Einsatz von Algorithmen nicht Pflicht sein. Robbins ist zudem der Meinung, dass Algorithmen überflüssig wären, wenn man fordert, dass sie erklärbar sein müssen. Denn für eine nützliche Erklärung eines Algorithmus müsse man bewerten können, welche Erwägungen akzeptabel sind und welche nicht – also zum Beispiel woran festgemacht wird, ob ein Muttermal bösartig ist. Wenn man aber schon wüsste, welche Erwägungen akzeptabel sind und an welchen

883 Freiesleben, König 2023:12
884 Ebd.:3–12

Kriterien man dementsprechend feststellen kann, ob ein Muttermal bösartig ist, dann brauche man keinen Algorithmus mehr, der die Kriterien durch Mustererkennung selbst festlegt. Stattdessen könne man einem Wenn-Dann-Algorithmus die genauen Kriterien vorgeben.[885] An dieser Stelle widerspricht Buijsman: „(…) we often do seem to be capable of evaluating decisions even if we can't write down the exact steps needed for making or even evaluating the decision."[886] Wir wissen zum Beispiel, dass wir eine Katze sehen, aber sind trotzdem nicht dazu in der Lage, notwendige und hinreichende Bedingungen dafür anzugeben, wie eine Katze aussieht. Selbst wenn wir Erklärungen bewerten können und feststellen, ob der Algorithmus die richtigen Dinge betrachtet, um zu entscheiden, ob etwas eine Katze ist, bedeutet das nicht, dass wir stattdessen einen Wenn-Dann-Algorithmus spezifizieren können. Laut Buijsman kann deshalb Erklärbarkeit immer noch eine sinnvolle Anforderung an komplizierte Algorithmen sein.[887]

Als nächstes widmet sich Buijsman dem Argument des Technologieethikers Alex John London, der der Meinung ist, wir forderten von Algorithmen eine Erklärbarkeit, die Menschen für ihre Entscheidungen auch nicht bieten könnten. Die meisten menschlichen Entscheidungen seien nicht interpretierbar in dem Sinne, dass die Menschen ein Modell, das sie für die Entscheidungsfindung verwenden, simulieren könnten. Menschliche Entscheidungen seien oft in dem Sinne interpretierbar, dass wir sie im Nachhinein rationalisieren können, und das gebe nicht unbedingt Aufschluss darüber, warum eine Person die Entscheidung getroffen habe. In diesem Sinne seien auch Algorithmen interpretierbar. Aus Londons Sicht ist es dann sogar ein Vorteil von Algorithmen gegenüber menschlichen Entscheidungen, dass Verlässlichkeit und Genauigkeit der ersteren leicht bewertet werden können. Die beste Grundlage für die Entscheidung, welche Behandlung angeboten werden soll, sei die Wirksamkeit der Behandlungen. Daher sollten wir algorithmischen Ergebnissen aufgrund ihrer Wirksamkeit vertrauen und nicht aufgrund der Frage, ob wir sie genau kennen.[888]

885 Robbins 2019:495, 510ff.; Buijsman 2023:198
886 Buijsman 2023:198
887 Ebd.:199
888 London 2019:18ff.; Buijsman 2023:199f.

Buijsman wendet als erstes ein, dass Menschen sehr wohl erklären können, warum sie eine bestimmte Entscheidung getroffen haben. So können zum Beispiel Ärzt:innen, die eine Behandlung verordnen, dies mit früheren Erfahrungen mit Patient:innen mit ähnlichen Symptomen begründen. Algorithmen hingegen schätzen wir für ihre Zuverlässigkeit und nicht für ihre Fähigkeit, Entscheidungen zu begründen. Deshalb sei es wichtig, dass Algorithmen nur als Hilfsmittel verstanden werden, welche bei der Entscheidungsfindung unterstützen können. Buijsman schreibt über London: „(…) the cases he highlights show that making decisions for the wrong reasons (theoretical ones rather than evaluated efficacy) can be damaging, and so that having these reasons can be an important part of figuring out how good a decision is."[889] Aus Buijsmans Sicht liefert London damit unabsichtlich einen Grund, erklärbare Algorithmen zu bevorzugen, damit gezeigt werden kann, dass deren Entscheidungen auf den relevanten Merkmalen, wie zum Beispiel der Wirksamkeit einer empfohlenen Behandlung, beruhen.[890]

Der Einwand gegen die Erklärbarkeit von Algorithmen von Juan Manuel Durán und der Bioethikerin Karin Rolanda Jongsma lautet, dass undurchsichtige Algorithmen nur mit Hilfe von weiteren undurchsichtigen Algorithmen erklärt werden können: „The fundamental problem with transparency is, to our mind, that it is itself based on opaque processes."[891] Dadurch, dass ein weiterer Algorithmus angewendet wird, um den ursprünglichen, undurchsichtigen Algorithmus erklärbar zu machen, wird die Frage nach den Entscheidungskriterien nur auf den neuen, erklärenden Algorithmus verlagert. Dabei können wir wissen, wie dieser zweite Algorithmus Erklärungen generiert, aber wir können weiterhin nicht sicher sein, ob die Erklärungen des zweiten Algorithmus tatsächlich korrekte Beschreibungen dafür sind, warum der ursprüngliche Algorithmus zu seinen Ergebnissen gekommen ist.[892]

Buijsman stimmt zu, dass wir undurchsichtige Algorithmen nicht loswerden könnten, wenn alle Erklärungswerkzeuge selbst undurchsichtige Algorithmen wären, aber er belegt, dass das nicht der Fall

889 Buijsman 2023:200
890 Ebd.:200
891 Durán, Jongsma 2021:331
892 Ebd.:330f.; Buijsman 2023:201

ist. Denn Methoden der Merkmalsbedeutung (feature importance methods), wie zum Beispiel die Datenvisualisierungstechnik der Heatmaps, machen einen Teil des Verhaltens des Algorithmus transparent und kontrafaktische Erklärungen zeigen transparent auf, welche minimalen Änderungen zu einer Änderung der Ergebnisse führen. Auch wenn es noch keine Tools gibt, die Black-Box-Algorithmen vollständig erklären können, gibt es doch Erklärungsansätze, die das Vorgehen von Algorithmen transparenter machen.[893]

Zum Schluss widmet sich Buijsman dem Argument von Krishnan, welches besagt, dass Erklärungen weiteren Zielen dienen, wie dem Schutz vor Diskriminierung, und diese Ziele auch durch andere Mittel erreicht werden können. Sie argumentiert, dass die Interpretierbarkeit oft nicht um ihrer selbst willen wichtig ist, sondern sie dient der Erreichung eines weiteren Ziels, wie z. B. Sicherheit, Gewährleistung der Genauigkeit und Nichtdiskriminierung: „(…) since interpretability is most often proposed as a means to further ends rather than an end in itself, it would be more perspicuous to organize discussion around the fundamental problems rather than one putative solution."[894] Krishnan schlägt vor, dass anstatt über die Interpretierbarkeit zu diskutieren, über die Ziele der Interpretierbarkeit gesprochen werden sollte. Dadurch könnte nicht mehr angenommen werden, dass es nur einen Weg gibt, ein bestimmtes Ziel zu erreichen und Aufmerksamkeit sowie Ressourcen würden nicht mehr ausschließlich auf eine bestimmte Art von Lösung konzentriert.[895]

So werde zum Beispiel oft angenommen, dass die Interpretierbarkeit eine wichtige Rolle bei der Rechtfertigung von algorithmischen Ergebnissen spielt. Es erscheint geradezu als unverantwortlich, den Ergebnissen von solchen Algorithmen zu glauben, die auf unbekannten realen Daten basieren, wenn keine detaillierten Kenntnisse über die Funktionsweise des Algorithmus vorliegen. In manchen Kontexten gibt es jedoch laut Krishnan Möglichkeiten, die gewünschte Sicherheit zu erreichen, auch wenn man das Innenleben der Werkzeuge nicht kennt: „(...) there are ways of being justifiably confident in outputs that circumvent the call for interpre-

893 Buijsman 2023:202
894 Krishnan 2020:500
895 Ebd.:495

tation (...)."[896] Die Position der Reliabilisten liefere eine plausible Erklärung dafür, warum man sich auf die Ergebnisse des visuellen Systems oder des Gehirns verlassen kann, auch wenn es keine detaillierte wissenschaftliche Theorie gibt: weil das System zuverlässig das richtige Ergebnis produziere. In manchen Fällen könne eine Erfolgsbilanz eine genaue Beschreibung der inneren Funktionsweise eines Instruments ersetzen.[897]

Wenn die Ergebnisse diskriminierend sind, weil ein Gesichtserkennungs-Algorithmus zum Beispiel Menschen mit dunkler Hautfarbe als Gorillas identifiziert[898], dann ist das Problem laut Krishnan auch ohne detailliertes Wissen über das Vorgehen des Algorithmus diagnostizierbar. In diesem Fall waren die Ergebnisse eindeutig diskriminierend und als festgestellt wurde, dass in den Trainingsdaten keine Bilder von Menschen mit dunkler Hautfarbe enthalten waren, wurde auch klar, warum. Laut Krishnan ist es nicht notwendig, die Details des algorithmischen Vorgehens zu verstehen, um überprüfen zu können, ob der Algorithmus diskriminierende Ergebnisse erzeugt. Stattdessen könnten bessere Verfahren zur Überprüfung der Konstruktion von Trainingsdatensätzen oder zum Testen von Entscheidungskriterien anhand von Beispieldatensätzen entwickelt werden, um zu überprüfen, wie der Algorithmus Personen unterschiedlicher Identitätskategorien behandelt.[899] Der Wunsch nach Interpretierbarkeit wurde zwar durch die Notwendigkeit motiviert, Diskriminierung zu bekämpfen[900], aber es gibt keine offensichtliche Verbindung zwischen der Bekämpfung von Diskriminierung und dem Wissen über die innere Funktionsweise von Algorithmen. Vielmehr kann die Interpretierbarkeit als ein Werkzeug unter vielen verstanden werden, das sich als nützlich gegen Diskriminierung erweisen kann.[901] Allerdings lenkt Krishnan ein, dass es auch Kontexte gibt, in denen die Interpretierbarkeit des Algorithmus als Selbstzweck wichtig ist. So zum Beispiel, wenn Algorithmen in der

896 Krishnan 2020:496
897 Ebd.:495f.
898 Guynn 2015; Barr 2015
899 Krishnan 2020:496, 498
900 Goodman, Flaxman 2017:55; Krishnan 2020:497
901 Krishnan 2020:497

Wissenschaft eingesetzt werden, um kausal relevante Faktoren und ihre Wechselwirkungen zu identifizieren.[902]

Buijsman hat sich von Krishnan davon überzeugen lassen, dass Erklärbarkeit kein Allheilmittel ist. Er gesteht ein, dass es gerechtfertigt sein kann, algorithmischen Ergebnissen zu glauben, ohne vollständige Erklärung oder Interpretation, weil man stattdessen zum Beispiel die Trainingsdatensätze überprüfen kann. Trotzdem hält Buijsman eine Erklärung der Funktionsweise des Algorithmus für sinnvoll, um mit Unstimmigkeiten umgehen und Beweise aus verschiedenen Quellen miteinander in Einklang bringen zu können. Eine Erklärung könne außerdem bei der Fehlersuche helfen und dazu beitragen, die algorithmischen Ergebnisse anzufechten und Verantwortung für Fehler einzufordern.[903] Die Erklärbarkeit kann also einerseits einen instrumentellen Wert haben, da sie es den Datennutzenden ermöglicht, Algorithmen zu verwalten.[904]

Andererseits ist die Erklärbarkeit von algorithmischen Ergebnissen dem Datenethiker Nathan Colaner zufolge auch intrinsisch wertvoll, weil nicht erklärbare Systeme entmenschlichend sein können. Zum einen geht es Colaner dabei um die Möglichkeit der Partizipation am Entscheidungsprozess: „(…) personal dignity requires the opportunity for meaningful participation in the process when the outcome affects me or my community, and meaningful participation is difficult or impossible if the algorithmic process is fundamentally un-understandable, regardless of whether the outcome is morally and legally acceptable."[905] Bei einem unverständlichen Prozess, dessen Ergebnis eine Person betrifft, könne diese Person nicht oder nur schwer sinnvoll an diesem Prozess beteiligt werden, und das verletze ihre persönliche Würde. Zum anderen erfordere die persönliche Würde eine Möglichkeit zu verstehen, warum ein Ereignis eingetreten ist, und das sei ebenfalls schwer zu erreichen, wenn der algorithmische Prozess grundsätzlich unverständlich ist.[906] Drittens erfordert die vollständige Verwirklichung des Menschen laut Colaner, dass wir in bestimmten Fällen die Verantwortung für

902 Krishnan 2020:499
903 Buijsman 2023:205
904 Colaner 2022:231f.
905 Ebd.:235
906 Ebd.:236

unser eigenes Leben behalten, unabhängig davon, ob ein Algorithmus dies effizienter tun könnte.[907]

Ich möchte dem hinzufügen, dass algorithmische Ergebnisse auch diskriminierend sein können, ohne dass Datennutzende dies bemerken, weil bei einem Bewerbungsprozess zum Beispiel nicht alle Bewerber:innen eingeladen werden können. Wenn die Ergebnisse unbemerkt diskriminierend sind, ist das Problem aus meiner Sicht nicht ohne Wissen über die Trainingsdaten und über die Gewichtung verschiedener Entscheidungskriterien durch den Algorithmus diagnostizierbar. Man braucht vielleicht keine detaillierten Informationen über die Funktionsweise des Algorithmus, aber ich habe argumentiert, dass man eine Antwort auf die Frage braucht, warum diese Ergebnisse erzeugt wurden, um Beweise sammeln zu können für die Verlässlichkeit von algorithmischen Ergebnissen.

7.2.2 Definition von Transparenz im Kontext von Algorithmen

In Bezug auf Algorithmen bedeutet Transparenz, dass zugänglich und verständlich sein muss, wie das Ergebnis zustande gekommen ist. Man könnte also meinen, es müssten technische Details über die Funktionsweise der Algorithmen offenbart werden.[908] Es stellt sich also die Frage, was Transparenz in Bezug auf Algorithmen heißen soll, wenn es nicht möglich ist, alle technischen Details offenzulegen und das zudem auch nicht erstrebenswert erscheint, weil es sehr schwer wäre diese technischen Details zu verstehen. Deshalb muss gefragt werden, was eine solche Erklärung für algorithmische Ergebnisse leisten soll. Sie soll verhindern, dass Datennutzende verzerrten Ergebnissen ausgesetzt werden. Sie soll also Fehler aufdecken. Ich argumentiere, dass Fehler in algorithmischen Ergebnissen aufgedeckt werden können, wenn klar ist, warum dieses Ergebnis erzeugt wurde.

Die Frage, warum ein Algorithmus ein bestimmtes Ergebnis erzeugt hat, kann laut Krishnan auf unterschiedliche Weise verstanden werden: im kausalen oder im rechtfertigenden Sinne. Wir fragen nach dem Warum im kausalen Sinne, wenn wir fragen, wie es dazu

907 Colaner 2022:237
908 Tutt 2017:111; Herzog 2022:4

kam, dass der Algorithmus das Ergebnis erzeugte, das er erzeugte. Wir fragen jedoch nach dem Warum im rechtfertigenden Sinne, wenn wir fragen, welche Gründe für das Ergebnis als Antwort auf eine bestimmte Frage sprechen.[909] Wenn zum Beispiel der von Amazon im Bewerbungsprozess eingesetzte Algorithmus eine Bewerberin ablehnt, kam es zu dem Ergebnis wegen ihres Geschlechts (kausal), und der Grund für das Ergebnis ist, dass in der Vergangenheit vorrangig Männer erfolgreich waren im Bewerbungsprozess und der Algorithmus dies reproduziert (rechtfertigend). Hier zeigt sich, dass sich die kausale Beschreibung eines algorithmischen Ergebnisses nicht ohne Weiteres auf rechtfertigende Überlegungen übertragen lässt. Denn selbst wenn man weiß, was den Algorithmus dazu veranlasst hat, eine Ablehnung zu empfehlen, heißt das noch nicht, dass man auch den Grund für die Ablehnung angeben kann.

Deshalb bezweifelt Krishnan, dass eine Interpretation – im Sinne von Informationen über den Prozess des Algorithmus – hilfreich ist für das Verständnis der algorithmischen Ergebnisse.[910] Ich hingegen ziehe aus dieser Unterscheidung zwischen dem kausalen und dem rechtfertigenden Warum die Erkenntnis, dass man zum Verständnis algorithmischer Ergebnisse nicht nur die Ergebnisse und ihren Entstehungsprozess betrachten muss, sondern immer auch die ihnen zugrundeliegenden Datensätze (Input). Auf diese Weise können Verzerrungen identifiziert werden, die sich bei der Datenverarbeitung in den Ergebnissen reproduzieren würden. So kann geprüft werden, ob die Daten die Wirklichkeit abbilden und ob es sich dabei um eine ungerechte Wirklichkeit handelt, die nicht reproduziert werden sollte. Die Analyse der algorithmischen zugrundeliegenden Datensätzen ist somit auch ein wichtiger Schritt, damit bereits bestehende Vorurteile nicht in Form von einer bestehenden Verzerrung (preexisting bias) in das System einfließen.[911]

In Bezug auf Algorithmen bedeutet Transparenz also weiterhin, dass zugänglich und verständlich sein muss, wie das Ergebnis zustande gekommen ist.[912] Post-hoc Erklärungen von Algorithmen sollen zu diesem Zweck aufzeigen, warum der Algorithmus eine

909 Krishnan 2020:493
910 Ebd.:494
911 Friedman, Kahn 2008:1253; Friedman, Nissenbaum 1996:332–335
912 Mittelstadt et al. 2016:6

bestimmte Ausgabe erzeugt. Sie tun das, indem das Verhältnis zwischen Input und Output offengelegt wird. Dadurch ist es möglich, die Frage nach dem Warum im kausalen Sinne zu beantworten. Um die Frage nach dem Warum im rechtfertigenden Sinne beantworten zu können, müssen die dem zu erklärenden Algorithmus zugrundeliegenden Datensätze betrachtet werden – denn nur diese bieten Aufschluss darüber, ob Verzerrungen vorliegen und welche Aussagen eigentlich gemacht werden können.[913] Durch diese Transparenz der Ergebnisse in Form von Erklärungen der Zusammenhänge zwischen Input und Output, sowie der Analyse der zugrundeliegenden Datensätze, können Verzerrungen und Fehler in den Ergebnissen aufgedeckt werden und somit kann Diskriminierung auf Grund bestimmter Merkmale im Input verhindert werden. Ich schließe, dass man hierdurch vollständig versteht, wie das System seinen Output erreicht, weil man alle epistemisch relevanten Elemente des Prozesses kennt, wenn man den Input kennt und weiß, welchen Output er erzeugt – so reduziert sich die epistemische Undurchsichtigkeit.[914]

7.2.3 Einbettung in den Kontext als Lösung

Trotz aller Widrigkeiten halte ich diese Transparenz im Kontext von Algorithmen für umsetzbar, indem die Ergebnisse im zum Verständnis notwendigen Kontext dargestellt werden. In diesem Teilkapitel wird erörtert, warum das die beste Lösung ist.

Man könnte meinen, wie unter 7.1 angedeutet, dass es zum Schutz der Datennutzenden und Vorbeugen von Verzerrungen sinnvoll wäre, sensible Daten von vorneherein nicht in Trainingsdatensätze oder Datenabfragen zu integrieren. Im Falle einer Berufsempfehlung und auch beim Wohngeldantrag sollten die Ergebnisse aus ethischer Sicht nicht abhängig sein von dem Geschlecht oder der Herkunft der Datenquellen und -nutzenden – hier wäre es durchaus möglich, solche sensiblen Daten gar nicht erst abzufragen und auch aus den Trainingsdatensätzen zu entfernen. Aber in anderen Fällen, zum Beispiel im medizinischen Bereich, können sensible Daten, wie das Geschlecht, ausschlaggebend dafür sein, wie hilfreich das algorith-

913 Mehr dazu im nächsten Teilkapitel unter 7.2.3.
914 Humphreys 2009:618; Buijsman, Veluwenkamp 2023:545

mische Ergebnis ist. In der Vergangenheit wurden hier solche Merkmale sogar zu wenig beachtet, sodass erst neueste Erkenntnisse auf Unterschiede in Krankheitsverläufen sowie effektiven Behandlungsmöglichkeiten zwischen Frauen und Männern hindeuten.[915] Es ist also nicht in allen Fällen ratsam, sensible Daten, wie das Geschlecht, außen vor zu lassen – auch weil man mit wenigen unspezifischen Daten nur unspezifische Ergebnisse erwarten kann und das dem Zweck des Einsatzes der Algorithmen zu widersprechen scheint.

Wenn Datenverarbeitende die Verzerrung aus den Trainingsdaten nicht rausrechnen können und verzerrte Ergebnisse auch anderweitig nicht vermeiden können, haben sie die Pflicht, Datennutzende darüber aufzuklären. So könnte man im Fall einer Berufsempfehlung zum Beispiel die Datennutzenden darüber aufklären, dass in der Vergangenheit vor allem Männer berufstätig waren, und die Trainingsdaten, auf welchen der Algorithmus basiert, deshalb 80 % Männer und 20 % Frauen repräsentieren. Die Datennutzenden, die sich durch diese Trainingsdaten nicht oder nur schlecht repräsentiert fühlen, haben daraufhin eine Erklärung dafür und einen Anlass, die Ergebnisse des Algorithmus zu hinterfragen. Insbesondere sollte durch eine solche Aufklärung für die Datennutzenden deutlich werden, was der Algorithmus leisten kann und was nicht. Denn ein Algorithmus, der auf beruflichen Werdegängen von Menschen in der Vergangenheit beruht, kann nicht vorhersagen, welcher berufliche Weg der Beste sein wird für eine Person heute. Erstens verändern sich Ausbildungswege und ganze Berufsfelder durch den Einsatz von neuen Technologien. Zweitens sagt zum Beispiel die Anzahl von Frauen in bestimmten Berufsgruppen nichts darüber aus, wie viele Frauen in Zukunft in diesen Berufsgruppen arbeiten sollten – das wäre ein naturalistischer Fehlschluss.[916]

Ein solcher Algorithmus kann nur als Anstoß dienen für diejenigen, die noch gar keine Idee haben, in welche Richtung es beruflich gehen könnte und welche Optionen es eigentlich gibt. Selbst dann kann ein Algorithmus, der auf beruflichen Werdegängen von Menschen in der Vergangenheit beruht, nur aussagen „Angesichts der Tatsache, dass in der Vergangenheit die meisten Menschen mit deinen Voraussetzungen folgende Berufe ergriffen haben, ist es emp-

915 Robert Koch-Institut 2020
916 Hume 2000[1739]:Buch III, Teil I, Kapitel I

fehlenswert, dass du auch einen dieser Berufe ergreifst". Es hängt dann von den Trainingsdaten ab, ob man überhaupt eine Aussage darüber machen kann, ob die Menschen in der Vergangenheit zufrieden waren mit ihren beruflichen Entscheidungen. Wenn es dazu keine Informationen gibt, ist es nicht einmal möglich, das Kopieren des Verhaltens anderer Menschen als Empfehlung zu bezeichnen – es handelt sich vielmehr um eine Aussage darüber, was in der Vergangenheit mit den gegenwärtigen Qualifikationen und Voraussetzungen möglich war. Durch solch eine transparente Darstellung der Leistung des Algorithmus und seiner Ergebnisse wird für Datennutzende klar, dass es auch andere Möglichkeiten gibt, als der Empfehlung zu folgen. So kann womöglich die epistemische Abhängigkeit vom Algorithmus verringert werden.[917]

Hier wird noch einmal deutlich, für wen die Ergebnisse transparent sein sollen, nämlich für die Datennutzenden. Aber das heißt nicht, dass jedem Datennutzenden die Struktur des verwendeten Algorithmus offenbart werden soll. Der Durchschnittsnutzende verfügt nicht über fundierte Kenntnisse der Informatik und hat möglicherweise nicht einmal ein starkes Interesse daran, diese zu erlernen.[918] Selbst wenn Datenverarbeitende betriebliche Informationen offenlegen, ist der Nutzen für die Gesellschaft deshalb laut Mittelstadt et al. ungewiss: „Transparency disclosures may prove more impactful if tailored towards trained third parties or regulators representing public interest as opposed to data subjects themselves (…)."[919] Auch die Sozioinformatikerin Katharina Zweig plädiert dafür, dass geschulte Dritte Einblick in das algorithmische Entscheidungssystem bekommen. Sie schlägt vor, dass dies ein ausgewählter Kreis von Expert:innen sein sollte – insbesondere wenn ein Algorithmus schwerwiegende Fehlurteile erzeugt hat. Aber wichtiger noch ist in ihren Augen, dass sowohl Trainingsdaten und -kriterien als auch Wirkungsweise und Entscheidungsregeln transparent dargestellt werden. Laut Zweig ist es immer sinnvoll, über Trainingsdaten und -kriterien solcher Algorithmen aufzuklären, die Menschen in Kategorien einteilen oder Urteile fällen über ihr zukünftiges Verhal-

917 Van den Hoven 1998:104f.; Rooksby 2009:82; Buijsman, Veluwenkamp 2023:546

918 Granka 2010:368

919 Mittelstadt et al. 2016:7

ten. Wenn aber die Wirkungsweise und Entscheidungsregeln des Algorithmus dargestellt werden sollen, schließe das die Methoden des maschinellen Lernens aus, deren Entscheidungsregeln selbst für die Programmierenden nicht mehr nachvollziehbar sind.[920]

Denn in solchen Algorithmen können mehr Merkmale aufgenommen werden als Menschen erfassen können, sodass ein Verständnis der Algorithmen, samt Fehlersuche und Verbesserung, sehr schwierig ist. Laut Burrell kann mit dieser Art von Undurchsichtigkeit umgegangen werden, indem Algorithmen vereinfacht werden. So ist es zum Beispiel möglich zu analysieren, welche Merkmale für das Ergebnis wichtig sind, und alle anderen Merkmale aus dem Modell zu entfernen.[921] Die Vereinfachung von Algorithmen würde auch Zweigs Forderung nach Wirkungsweise und Entscheidungsregeln des Algorithmus ermöglichen. Auch Mittelstadt et al. halten dies für hilfreich zum Verständnis: „Releasing information about the algorithm's decision-making logic in a simplified format can help (…)."[922] Wenn aber das Vorgehen der Algorithmen heruntergebrochen wird auf eine für den Menschen überschaubare Liste von Schlüsselkriterien, wird laut Burrell nur ein unvollständiges Verständnis vermittelt.[923]

Die bereits erwähnte KI-Ethikerin Kate Vredenburgh stimmt Burrell zu und wendet ein, dass es nicht ausreicht, vereinfachte Modelle von komplexen Algorithmen bereitzustellen, da es hierdurch zu irreführenden Erklärungen kommen kann und somit nur eine Illusion des Verstehens erzeugt wird.[924] Der Technologieethiker Christian Herzog merkt an, dass interpretierbare Algorithmen meist nur für ihre Entwickler:innen verständlich sind, bei einem solchen interpretierbaren Modell dann aber die Bemühungen fehlen, die Ergebnisse anderen Akteuren zu erklären.[925] Munch et al. geben zudem zu bedenken, dass es potenziell kognitiv nicht möglich ist, Verständnis von algorithmischen Ergebnissen zu erlangen, indem sie individuell mit Hilfe von XAI-Werkzeugen erklärt werden.[926] Der Soziologe

920 Zweig 2019:9
921 Burrell 2016:9
922 Mittelstadt et al. 2016:7
923 Burrell 2016:9
924 Vredenburgh 2022:225
925 Herzog 2021:222
926 Munch et al. 2024:9

Daan Kolkman bestätigt diese Bedenken, da er bei Fallstudien mit Algorithmusexpert:innen festgestellt hat, dass selbst vergleichsweise einfache Modelle sogar für diejenigen, die täglich mit ihnen arbeiten, schwer zu verstehen sein können.[927] Insbesondere für Menschen, die nicht täglich mit Algorithmen arbeiten, ist deshalb das Verständnis ihres Vorgehens und damit der Ergebnisse eine Herausforderung.[928]

Während Vredenburgh als Lösung vorschlägt, dass Expert:innen kostenlos bereitgestellt werden sollten[929], schlägt Kolkman vor, dass es eine bessere Kommunikation über Algorithmen für die breite Öffentlichkeit geben sollte[930]. Die Ansätze beider Autor:innen ergänzen sich, denn man braucht Expert:innen für die Erklärung der Algorithmen und anschließend eine gute Kommunikationsstrategie, um diese Erklärung der breiten Öffentlichkeit verständlich zu machen. Nimmt man noch die Lösung von Munch et al. hinzu, bräuchte man sogar mehrere Kommunikationsstrategien, da sie vorschlagen, Erklärungen auf verschiedene Personengruppen zuzuschneiden.[931] Die Hoffnung ist, dass es durch eine solche transparente Darstellung für Datennutzende möglich ist, alle epistemisch relevanten Elemente des Prozesses zu kennen.

Buijsman und der Digitalethiker Herman Veluwenkamp behaupten jedoch, dass es auf Grund der epistemischen Abhängigkeit nicht einmal reicht, alle epistemisch relevanten Elemente des Prozesses zu kennen, um epistemisch unabhängig zu sein: „It is possible to know all the epistemically relevant elements of the process (i.e. fully understand how the system reaches its output) while still being dependent on the system, because of a lack of outside information."[932] Wenn alternative Informationsquellen fehlen, sind die Datennutzenden auch dann noch epistemisch abhängig von den algorithmischen Ergebnissen, wenn sie deren Entstehung nachvollziehen können. Nun erscheint es zu viel verlangt, dass Datenverarbeitende neben den algorithmischen Ergebnissen auch noch alternative Informationsquellen bereitstellen.

927 Kolkman 2022:105f.
928 Ebd.:106
929 Vredenburgh 2022:225
930 Kolkman 2022:106
931 Munch et al. 2024:9
932 Buijsman, Veluwenkamp 2023:545

Deshalb schlage ich vor, dass die Ergebnisse im zum Verständnis notwendigen Kontext dargestellt werden sollten. So kann zum Beispiel das Ergebnis eines Berufsempfehlungs-Algorithmus, der auf Trainingsdaten aus der Vergangenheit basiert, durch die Information eingeordnet werden, dass der Algorithmus nur reproduzieren kann, was in der Vergangenheit mit den gegenwärtigen Qualifikationen und Voraussetzungen möglich war. Dafür ist es natürlich notwendig, dass die Datenverarbeitenden nachvollziehen können, wie das Ergebnis zustande gekommen ist, bzw. welcher Input zu welchem Output führt. Es muss für Datennutzende klar sein, dass es sich bei algorithmischen Ergebnissen nicht um objektiv gesetztes Wissen handelt. Diese Tatsache erfordert einen reflektierten Umgang mit algorithmischen Ergebnissen auf Seiten der Datennutzenden. Dieser sollte unterstützt werden durch die Datenverarbeitenden, indem sie den Kontext der Ergebnisse möglichst verständlich einordnen.

7.3 Moralische Pflicht zur Transparenz und daraus abgeleitetes Recht auf Transparenz

Eine Pflicht zur Transparenz sowie ein daraus abgeleitetes Recht auf Transparenz würden vor Handlungen schützen, die das Risiko einer erheblichen Beeinträchtigung der Transparenz mit sich bringen. Eine solche moralische Pflicht und das daraus resultierende Recht auf Transparenz schützen also vor Handlungen, die Verzerrungen und Fehler in den algorithmischen Ergebnissen verdecken und somit Diskriminierung schwer erkennbar machen. Das Ziel dieses Kapitels ist es, eine moralische Pflicht zur transparenten Darstellung algorithmischer Ergebnisse durch Datenverarbeitende zu begründen und zu verstehen, wie daraus für Datennutzende ein Recht auf Transparenz resultieren kann.

Solch eine moralische Pflicht lässt sich aus Kants erster Formulierung des Kategorischen Imperativs ableiten.[933] Denn wenn eine Handlungsmaxime weder gedacht noch gewollt werden kann, ist es eine vollkommene Pflicht, diese Handlung zu vermeiden. Eine Handlungsmaxime, die zwar denkbar ist, aber nicht widerspruchs-

933 Kant 2008:53 [421]

frei gewollt werden kann, etabliert eine unvollkommene Pflicht.[934] Es ist zwar denkbar, dass algorithmische Ergebnisse intransparent zur Verfügung gestellt werden, aber es kann nicht gewollt werden – das etabliert eine unvollkommene Pflicht zum Unterlassen der Bereitstellung intransparenter Ergebnisse. Diese Pflicht sowie eine damit einhergehende Pflicht zur Bereitstellung transparenter Ergebnisse lassen sich begründen, indem aufgezeigt wird, warum die fehlende Transparenz vermieden werden sollte und warum eine transparente Darstellung algorithmischer Ergebnisse erstrebenswert ist.

Vredenburgh argumentiert, dass ohne Erklärungen von Entscheidungen, die ein Individuum betreffen, die Handlungsfähigkeit des Individuums verminderte wäre – insbesondere würde es dem Individuum erschwert, sich für sich selbst einzusetzen: „The right to explanation is grounded in the interest in informed self-advocacy."[935] Diese informierte Selbstvertretung umfasst diverse Fähigkeiten, die man braucht, um die eigenen Interessen und Werte zu vertreten. Eine Person, die sich selbst vertritt, kann Regelsysteme so steuern, dass die eigenen Ziele erreicht und Entscheidungstragende für Fehler oder Ungerechtigkeiten zur Rechenschaft gezogen werden. Um sich selbst vertreten zu können, brauchen Individuen epistemische Unterstützung. Laut Vredenburgh würde durch eine kausale Erklärung einer eine Person betreffende Entscheidung klar, was das Individuum tun müsste, um ein von ihm gewünschtes Ergebnis zu erzielen.[936] Das Konzept der informierten Selbstvertretung betont die Bedeutung von Erklärungen für algorithmische Entscheidungen, die uns direkt betreffen, denn ohne solche Erläuterungen sind wir nicht in der Lage, unser Verhalten als Reaktion auf das algorithmische Ergebnis so anzupassen, dass unsere Interessen und Werte gefördert werden.[937]

---

934 Kant 2008:55-58 [422–424]

935 Vredenburgh 2022:212

936 Ebd.:213, 223f., 217

937 Munch et al. 2024:4f.; Vredenburgh 2022; Vredenburgh begründet mit diesem Argument ein Recht auf Erklärung, aber die Argumente für eine durch eine Erklärung transparente und gegen eine intransparente Darstellung algorithmischer Ergebnisse können ebenfalls zu der Schlussfolgerung führen, dass es eine Pflicht zur Transparenz geben sollte. Da in dieser Arbeit davon ausgegangen wird, dass ein Recht auf Transparenz nur aus einer moralischen Pflicht zur Transparenz resultieren kann (Mohr 2010:77), ist hier die Schlussfolgerung,

Ohne das Verständnis, warum ein Ergebnis zustande gekommen ist, können Datennutzende auch laut London nicht reflektiert Stellung dazu nehmen und das Ergebnis nicht anfechten: „(…) understanding why a decision was made enables stakeholders to evaluate its merits and hold experts accountable for avoidable error."[938] Durch die Bereitstellung von intransparenten Ergebnissen, können die Datennutzenden diese nicht einordnen und deshalb nicht reflektiert Stellung dazu nehmen und ihre eigenen Interessen nicht vertreten. Wenn man Datennutzende mit Ergebnissen konfrontiert, die sie einfach hinnehmen sollen, würdigt man nicht ihre Rationalität, sondern gebraucht sie sogar als bloßes Mittel. Denn ohne das Verständnis der Ergebnisse zu ermöglichen, unterbindet man ihre freiwillige und informierte Zustimmung. Das heißt fehlende Transparenz sollte in Bezug auf algorithmische Ergebnisse vermieden werden, weil Datennutzende dadurch als bloßes Mittel gebraucht werden und das widerspricht Kants dritter Formel des Kategorischen Imperativs.[939]

Der Philosoph Keith Harris bemerkt passend dazu, dass die Beschränkung des Zugangs zu Beweisen bereits als epistemische Dominanz (epistemic domination) gilt.[940] Diese Beziehung epistemischer Dominanz kann den Datennutzenden epistemisch zugutekommen, wenn die Datenverarbeitenden wohlmeinend und gut informiert sind. Wenn aber die Datenverarbeitenden schlechte Absichten haben, ist es wahrscheinlich, dass die Datennutzenden zu falschen Überzeugungen gelangen und womöglich die Orientierung verlieren, da sie nicht mehr wissen, was sie glauben können.[941] Da es für das menschliche Wohlergehen wichtig ist, dass wir uns von richtigen und nicht von falschen Annahmen leiten lassen[942], sollten Datenverarbeitende dafür sorgen, dass möglichst keine falschen Überzeugungen entstehen.

Eine transparente Darstellung algorithmischer Ergebnisse ist hingegen erstrebenswert, weil sich die Datennutzenden dadurch des Risikos bewusst werden, dem sie ausgesetzt sind, und idealerweise

---

dass es eine Pflicht zur transparenten Darstellung algorithmischer Ergebnisse geben sollte.

938 London 2019:16
939 Kant 2008:65 [429]
940 Harris 2022:136
941 Ebd.:137, 139
942 Alston 2005:30

auch die epistemische Asymmetrie ausgeglichen wird, sodass verzerrte Ergebnisse nicht mehr so leicht zu falschen Überzeugungen führen können. Wenn zum Beispiel ein Wohngeldantrag abgelehnt wird und diese Ablehnung für die Antragstellerin in der Rolle der Datennutzenden nicht nachvollziehbar ist, dann sollte es möglich sein herauszufinden, warum der Wohngeldantrag abgelehnt wurde. Denn die Datennutzenden sind auf Informationen über das Vorgehen des Algorithmus angewiesen, um darauf angemessen reagieren zu können.[943] Denn mit Hilfe von der transparenten Darstellung algorithmischer Ergebnisse können Datennutzende eine fundierte Entscheidung darüber treffen, wie sie mit den betreffenden Ergebnissen umgehen[944], und können auf Grundlage dieser Informationen ihre eigenen Interessen vertreten[945].

Auch wenn es bei transparenter Darstellung möglich ist, ist es schwierig für Datennutzende die Ergebnisse selbst überprüfen zu müssen. Es ist leichter für die Datenverarbeitenden, Algorithmen einzusetzen, die zu keinen verzerrten Ergebnissen führen. Hierbei können natürlich auch Fehler unterlaufen, aber wichtig ist, dass Datennutzende nicht bewusst falschen Ergebnissen ausgesetzt werden. Wenn Datenverarbeitende undurchsichtige Algorithmen programmieren, die auswählen, welche Ergebnisse Datennutzenden zugänglich sind, dann haben sie die Möglichkeit, die Datennutzenden einseitig zu informieren oder ihnen nur die Daten zur Verfügung zu stellen, die ein erwünschtes Verhalten generieren. Ein solches strategisches Vorenthalten von Informationen wäre ethisch problematisch, da die Datennutzenden dadurch manipuliert und instrumentalisiert würden. Es wäre aber durchaus möglich, Informationen über die algorithmischen Ergebnisse strategisch vorzuenthalten, da eine Informationsasymmetrie zwischen Datennutzenden und Datenverarbeitenden besteht.[946]

Eine Erklärung, im Sinne der Offenlegung des Verhältnisses zwischen Input und Output, sowie der Analyse des zugrundeliegenden Datensatzes, ermöglicht die transparente Darstellung der algorithmischen Ergebnisse – die Ergebnisse werden so im zum Verständnis notwendigen Kontext dargestellt. Erst diese transparente Darstellung

943 Munch et al. 2024:8
944 Ebd.:9
945 Vredenburgh 2022:212
946 Drew 1991:25

der Erklärung ermöglicht die informierte Selbstvertretung der Datennutzenden. Da die transparente Darstellung der Ergebnisse erstrebenswert ist, und fehlende Transparenz bei der Bereitstellung algorithmischer Ergebnisse nicht gewollt werden kann, haben Datenverarbeitende, welche öffentlich Informationen bereitstellen, eine Pflicht zur Transparenz.

Dabei sind laut Hare aus deontologischer Sicht genau die Handlungsfolgen relevant, welche die moralischen Grundsätze verbieten oder herbeiführen wollen.[947] Somit ist das Bereitstellen von intransparenten algorithmischen Ergebnissen unter anderem deshalb falsch, weil es die unbemerkte Diskriminierung bestimmter Gesellschaftsgruppen zur Folge haben kann. Der Grundsatz des Kategorischen Imperativs verbietet die Handlungsfolge der Diskriminierung, weil die zugrundeliegende Handlungsmaxime nicht gewollt werden kann und somit eine unvollkommene Pflicht entsteht, diese Handlung zu vermeiden. Entsprechend dieses Grundsatzes ist es moralisch relevant, wenn die Folge einer Handlung die Diskriminierung einer Gesellschaftsgruppe ist.[948]

Darum lässt sich die Pflicht zur Transparenz auch dadurch begründen, dass fehlende Transparenz dazu führen kann, dass bestimmte Gruppen auf Grund von Merkmalen benachteiligt werden, die eigentlich keine Rolle spielen sollten.[949] So wurden zum Beispiel von dem Algorithmus bei Amazon im Jahr 2014 Bewerberinnen auf Grund ihres Geschlechts nicht eingeladen[950], und jobsuchende Frauen haben im Jahr 2020 in Österreich auf Grund ihres Geschlechts mit geringerer Wahrscheinlichkeit Fördermaßnahmen finanziert bekommen[951]. Im Fall des Algorithmus zur Bewertung von Bewerbungen bei Amazon wäre bei transparenten Entscheidungskriterien schneller klar geworden, dass das KI-System die Bewerber:innen anhand ihres Geschlechts einordnet, und dass es sich bei dieser Unterscheidung um eine Verzerrung handelt. Denn es wurde weiblichen Bewerberinnen systematisch aus unvernünftigen und unangemessenen Gründen eine Chance vorenthalten, sodass

947 Hare 1993:15
948 Ebd.
949 Stahl et al. 2023:10f.
950 Ebd.:10f.
951 Wimmer 2019; Köver 2019; Pflügl 2024

sich ein ungerechtes Ergebnis eingestellt hat.[952] Das größte Problem dabei ist, dass solche Ergebnisse trotzdem als objektiv dargestellt werden: „(...) biased decisions are presented as the outcome of an objective algorithm."[953] Das schürt jedoch die falsche Erwartung, dass KI-Systeme Wissen im Sinne von wahren und gerechtfertigten Überzeugungen bereitstellen können – dabei handelt es sich nur um die Reproduktion von Mustern in den Trainingsdaten. Es kann nicht gewollt werden, dass algorithmische Ergebnisse verzerrt sein und deshalb zu Diskriminierung führen können, dies jedoch nicht offensichtlich ist, sodass die Datennutzenden algorithmische Ergebnisse nicht auf ihre Plausibilität überprüfen können und ihnen damit willkürlich ausgeliefert sind – insbesondere, wenn Entscheidungen von den oder über die Datennutzenden direkt von den Ergebnissen beeinflusst werden oder sogar abhängen. Deshalb sollten Datenverarbeitende beim Einsatz von Algorithmen und bei der Bereitstellung ihrer Ergebnisse eine Pflicht zur Transparenz haben.[954]

Außerdem ist laut Hare aus deontologischer Sicht das Bereitstellen von intransparenten algorithmischen Ergebnissen deshalb falsch[955], weil sich Datennutzende auf dieser Informationsgrundlage eine falsche Überzeugung bilden können und es schwierig ist, das Ergebnis anzufechten. Ohne die Möglichkeit dem Ergebnis zuzustimmen oder es abzulehnen, in Ermangelung an Informationen, werden die Datennutzenden als bloße Mittel gebraucht, die die Ergebnisse einfach nur akzeptieren sollen. Die dritte Formel des Kategorischen Imperativs verbietet aber, dass Datennutzende als bloßes Mittel gebraucht werden. Diese Konsequenz, Datennutzenden die Möglichkeit zu verwehren zu den Ergebnissen reflektiert Stellung zu nehmen (und sie dadurch bloß als Mittel zu gebrauchen), ist das, was das Bereitstellen intransparenter Ergebnisse aus deontologischer Sicht falsch macht.

Aus dieser vollkommenen moralischen Pflicht zur transparenten Darstellung algorithmischer Ergebnisse auf Seiten der Datenverarbeitenden kann ein Primärrecht auf Transparenz für Datennutzende resultieren.[956] In der Literatur ist die Ansicht weit verbreitet, dass

952 Friedman, Nissenbaum 1996:332
953 Goodman, Flaxman 2017:53
954 Tutt 2017:110, 116f; Mittelstadt et al. 2016:6
955 Hare 1993:15
956 Mohr 2010:77; Mieth, Bambauer 2016:3

Menschen ein Recht auf Erklärung haben, wenn sie algorithmischen Entscheidungen unterworfen sind, bei denen viel auf dem Spiel steht.[957] So argumentieren zum Beispiel die Epistemologen Jeremy Fantl und Matthew McGrath, dass die Tatsache, ob eine Überzeugung epistemisch gerechtfertigt ist, nicht nur von den Beweisen abhängt, die wir für die Wahrheit dieser Überzeugung haben, sondern zudem auch von pragmatischen Faktoren.[958] Der Gedanke findet sich auch bei Dormandy: „Könnte die falsche Handlung z. B. einen Todesfall verursachen, so benötigen wir eine viel bessere Rechtfertigung für die Überzeugungen, die unsere Handlungsbasis bilden, als wenn die schlimmste Konsequenz wäre, dass wir uns mit Vanilleeis statt mit Schokoladeneis zufriedengeben müssten."[959] Wenn also ein Algorithmus im medizinischen Bereich eingesetzt wird, und das Befolgen seiner Empfehlung zum Tod eines Patienten führen könnte, müsse man sich sicherer sein, dass der Glaube an die Empfehlung gerechtfertigt ist, als wenn ein Algorithmus empfiehlt, welche Kleidung man sich kaufen sollte.

Darauf erwidern jedoch Munch et al., dass auch algorithmische Entscheidungen mit geringem Risiko ein Recht auf Erklärung begründen können, wenn sie Teil eines Musters von Entscheidungen sind, deren Ergebnis insgesamt gesehen für den Einzelnen von erheblicher Bedeutung sein kann.[960] Ein Beispiel hierfür wäre die durch einen Algorithmus auf die Nutzenden maßgeschneiderte Werbung auf einer Social-Media-Plattform. Mit Hilfe des Algorithmus und Informationen aus den jeweiligen Profilen wird die Wahrscheinlichkeit optimiert, dass sich Nutzende mit der angezeigten Werbung auseinandersetzen. Die Nutzenden verstehen dabei nicht, nach welchen Kriterien der Algorithmus vorgeht, und durch die wiederholte Anzeige dieser gezielten Werbung über einen gewissen Zeitraum entwickeln die Nutzenden, eine Vorliebe für teuren Spezialitätenkaffee.[961] Einmaliger Kontakt mit der Kaffeewerbung hat wenig Einfluss auf das Verhalten der Nutzenden und birgt deshalb nur ein geringes Risiko. Wenn aber die Nutzenden die Werbung immer wieder wahr-

957 Munch et al. 2024:1; Herzog 2022:50
958 Fantl, McGrath 2002:88
959 Dormandy 2019:184
960 Munch et al. 2024:2
961 Ebd.:7

nehmen, kann das zu einer Veränderung ihres Verhaltens führen, was als erheblicher Effekt und damit größeres Risiko einzuschätzen ist.[962]

Die Nutzenden können nur dann eine fundierte Entscheidung darüber treffen, ob sie mit dem Algorithmus interagieren möchten, wenn sie wissen, wozu der Algorithmus eingesetzt wird. Ein Recht auf Erklärung besteht laut Munch et al. dann, wenn es sich um algorithmische Entscheidungen mit hohem Risiko handelt, oder wenn algorithmische Entscheidungen mit geringem Risiko Teil eines Musters von Entscheidungen sind, deren Ergebnis insgesamt von erheblicher Bedeutung sein kann.[963] Laut Munch et al. brauchen wir also Informationen über die gezielten Einwirkungen der Algorithmen auf Menschen, um bestimmen zu können, ob sie ein Recht auf eine Erklärung für einmalige algorithmische Entscheidungen mit geringem Risiko haben.[964] Im Sinne von Mohr ließe sich jedoch argumentieren, dass allein aus der moralischen Pflicht zur transparenten Darstellung algorithmischer Ergebnisse Primärrechte für Datennutzende resultieren, die in der Rechtsordnung verankert werden sollten.[965]

7.3.1 Rechtliche Lage im Vergleich zu ethischen Anforderungen

Tatsächlich ist zumindest die Transparenz von Informationen in Verwaltungsverfahren und algorithmischen Ergebnissen im deutschen und europäischen Recht institutionalisiert. Um zu überprüfen, inwiefern die ethische Pflicht zur Transparenz bei der Bereitstellung von öffentlichen Informationen schon in juridischen Rechten verankert ist, soll hier ein kleiner Überblick über die rechtliche Lage gegeben werden. Artikel 5, Absatz 1 des deutschen Grundgesetzes besagt: „Jeder hat das Recht, (...) sich aus allgemein zugänglichen Quellen ungehindert zu unterrichten."[966] Das heißt, staatliche Organe, wie die Bundesregierung, müssen aktiv Informationsarbeit leisten

962 Munch et al. 2024:8
963 Ebd.:2
964 Munch et al. 2024:10
965 Mohr 2010:77
966 Art. 5, Abs. 1, Satz 1 GG

und solche allgemein zugänglichen Quellen zur Verfügung stellen, damit Staatsbürger:innen die Politik beurteilen und bei demokratischen Wahlen informierte Entscheidungen treffen können. Zudem besagt das Gesetz zur Regelung des Zugangs zu Informationen des Bundes (Informationsfreiheitsgesetz – IFG) Folgendes: „Jeder hat nach Maßgabe dieses Gesetzes gegenüber den Behörden des Bundes einen Anspruch auf Zugang zu amtlichen Informationen. Für sonstige Bundesorgane und -einrichtungen gilt dieses Gesetz, soweit sie öffentlich-rechtliche Verwaltungsaufgaben wahrnehmen."[967] Wo also Verwaltungsaufgaben übernommen werden, haben Staatsbürger:innen ein Recht auf Zugang zu amtlichen Informationen – damit ist gemeint: „(...) jede amtlichen Zwecken dienende Aufzeichnung (...)."[968] Es gibt jedoch auch Ausnahmen zu diesem Recht, wenn es sich bei den amtlichen Informationen zum Beispiel um personenbezogene Daten von Dritten handelt. Dann darf der Zugang nur bewilligt werden, wenn die Dritten eingewilligt haben oder das Informationsinteresse das schutzwürdige Interesse der Dritten überwiegt.[969] Noch konkreter auf die öffentliche Verwaltung in Deutschland bezogen gibt es Informationsrechte, die im Verwaltungsverfahrensgesetz (VwVfG) festgehalten sind: „Die Behörde hat den Beteiligten Einsicht in die das Verfahren betreffenden Akten zu gestatten, soweit deren Kenntnis zur Geltendmachung oder Verteidigung ihrer rechtlichen Interessen erforderlich ist. Satz 1 gilt bis zum Abschluss des Verwaltungsverfahrens nicht für Entwürfe zu Entscheidungen sowie die Arbeiten zu ihrer unmittelbaren Vorbereitung."[970] Hier wird deutlich, dass zumindest der Zugänglichkeits-Teil[971] der ethischen Pflicht zur Transparenz bei der Bereitstellung von öffentlichen Informationen schon in juridischen Rechten verankert ist.

Es fragt sich nun, inwiefern sich diese rechtlichen Regelungen übertragen lassen, wenn in Verwaltungsverfahren Algorithmen eingesetzt werden. Beim Einsatz von Algorithmen könnte der Anspruch auf Zugang zu jeder „(...) amtlichen Zwecken dienende[n] Aufzeich-

967 §1, Abs. 1, Satz 1 und 2 IFG
968 §2, Abs. 1, Satz 1 IFG
969 §5, Abs. 1, Satz 1 IFG
970 §29, Abs. 1, Satz 1 und 2 VwVfG
971 Mittelstadt et al. 2016:6

nung (...)“[972] eine neue Bedeutung bekommen. Denn während Verwaltungsmitarbeitende vermutlich aufzeichnen, wie es zu gewissen Entscheidungen kommt, erzeugen Algorithmen nur ein Ergebnis. Angenommen, der Wohngeldantrag einer Person wird abgelehnt, und sie fordert Einsicht in die das Verfahren betreffenden Akten, dann hätte sie nach Abschluss des Verwaltungsverfahrens auch Anspruch auf Einsicht in Entwürfe zu Entscheidungen sowie deren Vorbereitung. Hat aber ein Algorithmus die Entscheidung getroffen bzw. wurde die Entscheidung von Verwaltungsmitarbeitenden allein auf der Grundlage der algorithmischen Empfehlung getroffen, gibt diese Information nicht viel Aufschluss darüber, warum es zu dieser Entscheidung kam.

Das deutsche Informationsrecht im Kontext der Verwaltung bezieht sich noch auf von Menschen durchgeführte und dementsprechend anhand von Akten nachvollziehbare Verwaltungsprozesse. Solche Prozesse würden durch den Einsatz von Algorithmen jedoch stark verändert, da das Vorgehen nicht mehr dokumentiert und womöglich nicht mehr nachvollziehbar wäre. An dieser Stelle kann das europäische Recht weiterhelfen, welches sich der digitalen Datenverarbeitung und ihren rechtlichen Umständen in den letzten Jahren ausführlich gewidmet hat. So enthält zum Beispiel die Datenschutzgrundverordnung (DSGVO) eine Informationspflicht bei der Erhebung von personenbezogenen Daten. Das heißt, um eine faire und transparente Verarbeitung zu gewährleisten, müssen bei einer automatisierten Entscheidungsfindung, von der Personen betroffen sind, unter anderem „(...) aussagekräftige Informationen über die involvierte Logik (...)“[973] bereitgestellt werden. Die involvierte Logik algorithmischer Entscheidungsfindung könnte durch Informationen über den Zusammenhang zwischen Input und Output dargestellt werden. Hier stimmen meine ethischen Forderungen mit den rechtlichen Vorschriften der DSGVO überein.

Außerdem haben laut DSGVO alle Personen das Recht, nicht einer Entscheidung unterworfen zu werden, die ausschließlich auf einer automatisierten Verarbeitung beruht.[974] Dies gilt allerdings nur dann, wenn die Entscheidung der Person gegenüber rechtliche Wir-

---

972 §2, Abs. 1, Satz 1 IFG
973 Art. 13, Absatz 2 f) und Artikel 14, Absatz 2 g) DSGVO
974 Art. 22, Abs. 1 DSGVO

kung entfaltet oder sie anderweitig erheblich beeinträchtigt, wie zum Beispiel eine „(...) automatische Ablehnung eines Online-Kreditantrags oder Online-Einstellungsverfahren ohne jegliches menschliche Eingreifen."[975] Bei der Verwendung eines hochkomplexen Algorithmus kann man nie ganz sicher sein, ob der Glaube an den Output gerechtfertigt ist. Deshalb schlägt Burrell vor, die Verwendung von Algorithmen in bestimmten kritischen Anwendungsbereichen zu vermeiden.[976] Der AI Act hingegen schließt nicht gesamte Anwendungsbereiche aus, sondern er verbietet bestimmte Praktiken im KI-Bereich.[977]

Der AI Act schreibt vor, dass KI-Systeme nach Intensität und Umfang ihrer Risiken bewertet und eingeteilt werden. Damit der Einsatz sogenannter Hochrisiko-KI-Systeme erlaubt ist, werden hohe Anforderungen gestellt und Transparenzpflichten festgelegt.[978] In Artikel 6 und den Anhängen I und III wird definiert, was ein Hochrisiko-KI-System ausmacht. Das hohe Risiko bezieht sich hier vor allem auf kritische Anwendungsbereiche und direkte Auswirkungen auf natürliche Personen.[979] Bestimmte KI-Systeme werden als hochriskant eingestuft, „(...) um nachteilige Auswirkungen zu vermeiden, das Vertrauen der Öffentlichkeit zu erhalten und die Rechenschaftspflicht und einen wirksamen Rechtsschutz zu gewährleisten."[980] An dieser Stelle geht die Ethik über das Recht hinaus, da sie ihre Vorsichtsmaßnahmen nicht auf KI-Systeme in hochriskanten Kontexten beschränkt. Denn laut Munch et al. können auch algorithmische Entscheidungen mit geringem Risiko ein Recht auf Erklärung begründen, und zwar wenn sie Teil eines Musters von Entscheidungen sind, deren Ergebnis insgesamt gesehen für den Einzelnen von erheblicher Bedeutung sein kann.[981]

Transparenz bedeutet im AI Act, dass KI-Systeme „(...) angemessen nachvollziehbar und erklärbar (...)"[982] sind, und dass den Nutzenden bewusst gemacht wird, dass sie mit einem KI-System

975 Erwägungsgrund 71 DSGVO
976 Burrell 2016:9
977 Art. 5 EU AI Act
978 Vorbemerkung Nr. 26 und 46 EU AI Act
979 Art. 6; Anhang I und III EU AI Act
980 Vorbemerkung Nr. 59 EU AI Act
981 Munch et al. 2024:2
982 Vorbemerkung Nr. 27 EU AI Act

interagieren.[983] Dafür müssen die Betreiber:innen die Nutzenden „(...) über die Fähigkeiten und Grenzen des KI-Systems informieren und die betroffenen Personen über ihre Rechte in Kenntnis setzen (...).“[984] Hochrisiko-KI-Systeme sollen transparent sein, damit die Ausgaben des Systems angemessen interpretiert und verwendet werden können.[985] Zum Erreichen der Transparenz sollen Hochrisiko-KI-Systeme mit Betriebsanleitungen ausgestattet werden, welche Informationen über ihre Merkmale, Fähigkeiten und Leistungsgrenzen enthalten. So soll deutlich werden, zu welchem Zweck das Hochrisiko-KI-System entwickelt wurde, aber auch, welche Risiken alle vorhersehbaren Fehlanwendungen bergen. Zudem soll die zu erwartende Genauigkeit des Hochrisiko-KI-Systems sowie seine Leistung in Bezug auf bestimmte Personen oder Personengruppen, auf die es angewandt werden soll, offenbart werden. Wenn sie zur Erläuterung der Ausgaben des Hochrisiko-KI-Systems relevant sind, müssen auch technische Merkmale und Informationen über die verwendeten Trainings-, Validierungs- und Testdatensätze bereitgestellt werden. Außerdem soll eine solche Betriebsanleitung Informationen beinhalten, „die es den Betreibern ermöglichen, die Ausgabe des Hochrisiko-KI-Systems zu interpretieren und es angemessen zu nutzen“.[986] Damit hierbei Geschäftsgeheimnisse gewahrt werden können, müssen die Informationen nicht technisch detailliert sein.[987]

Der in deutschen Gesetzen verankerte Zugang zu amtlichen Informationen sowie die Einsicht in die das Verfahren betreffenden Akten ermöglichen nur einen Teilaspekt der Transparenz: die Zugänglichkeit. Die Verständlichkeit der Informationen ist dadurch nicht garantiert.[988] Erst die europäischen Gesetze bzw. Verordnungen verpflichten zusätzlich dazu, die zur Verfügung gestellten Informationen oder algorithmischen Ergebnisse verständlich zu machen. Trotzdem leistet die Ethik hier einen wichtigen Beitrag, weil sie eine ganz genaue Analyse davon ermöglicht, welche Informationen in Bezug auf algorithmische Ergebnisse bereitgestellt werden sollten.

983 Art. 50, Abs. 1, Satz 1 EU AI Act
984 Vorbemerkung Nr. 27 EU AI Act
985 Art. 13, Abs. 1, Satz 1 EU AI Act
986 Art. 13, Abs. 3 (b) EU AI Act
987 Vorbemerkung Nr. 107 EU AI Act
988 Mittelstadt et al. 2016:6

Die rechtliche Pflicht, „(...) aussagekräftige Informationen über die involvierte Logik (...)“[989] von algorithmischer Entscheidungsfindung bereitzustellen, ist noch sehr unkonkret. Auch die Anforderungen an die Betriebsanleitung für Hochrisiko-KI-Systeme aus dem AI Act bleibt vage. Hier sollen die Informationen über die algorithmischen Ergebnisse ihre Interpretation und angemessene Nutzung ermöglichen.[990] Wann Informationen über die Logik eines Algorithmus aussagekräftig sind, beziehungsweise wann Informationen über algorithmische Ergebnisse deren Interpretation und angemessene Nutzung ermöglichen, wurde erst durch meine ethische Erläuterung spezifiziert. Mit Hilfe der Ethik ließ sich herausarbeiten, dass man wissen muss, welcher Input welchen Output erzeugt und welche Datensätze zugrunde liegen, um ein algorithmisches Ergebnis verstehen zu können[991], und dass die Ergebnisse im zum Verständnis notwendigen Kontext dargestellt werden sollten, um transparent zu sein. Wer dafür verantwortlich ist, diese Pflicht zu erfüllen, wird in Kapitel 8 erörtert.

7.4 Aus der Pflicht zur Transparenz abgeleitete Anforderungen

In diesem Kapitel ist klar geworden, dass Datenverarbeitende moralisch verpflichtet sind, Datennutzenden transparente Ergebnisse bereitzustellen, da sie ohne Transparenz nicht reflektiert Stellung nehmen können und somit als Mittel gebraucht werden. Die fehlende Transparenz kann auf Grund ihrer unerwünschten Folgen, von dem Bilden falscher Überzeugungen bis zur Unmöglichkeit die eigenen Interessen zu vertreten, als Risiko bezeichnet werden. Da dieses Risiko den Datennutzenden von den Datenverarbeitenden auferlegt wird, handelt es sich um eine Risikoauferlegung. Mit Hilfe der deontologischen Risikoethik lässt sich deshalb argumentieren, dass diese Risikoauferlegung nur im Rahmen eines gerechten sozialen Systems der Risikobereitschaft geschehen darf, welche allen Beteiligten zum

989 Art. 13, Absatz 2 f) DSGVO und Artikel 14, Absatz 2 g) DSGVO
990 Art. 13, Abs. 3 (b) EU AI Act
991 Humphreys 2009:618; Buijsman, Veluwenkamp 2023:545

Vorteil gereicht.[992] Das bedeutet, durch die Regel der dritten Formel des Kategorischen Imperativs von Kant, dass niemand als bloßes Mittel gebraucht werden darf.[993] Das heißt in der Praxis, das die individuelle Kontrolle der Datennutzenden über das Risiko sollte so groß wie möglich sein. Dafür sollten die Risikoexponierten ohne Täuschung über das Risiko informiert werden und ohne Zwang die Wahl haben, ob sie das Risiko tragen wollen. Auf diese Weise kann die Transparenz gefördert werden.

Aus den theoretischen Überlegungen zur Transparenz bezüglich algorithmischer Ergebnisse lassen sich Anforderungen an die tatsächliche Umsetzung von Datenverarbeitung ableiten. Definieren wir Transparenz, wie bei Mittelstadt et al., als Zugänglichkeit und Verständlichkeit[994], so müssen wir die Bedeutung algorithmischer Ergebnisse zugänglich und verständlich machen, indem wir ihre Entstehung betrachten – nur durch die Erklärung der Entstehung des algorithmischen Ergebnisses sind wir in der Lage, seine Bedeutung zu verstehen. Um den Datennutzenden das Risiko der fehlenden Transparenz im Rahmen eines gerechten sozialen Systems der Risikobereitschaft aufzulegen, sollten Datenverarbeitende ihrer Pflicht zur Transparenz nachkommen, indem sie die algorithmischen Ergebnisse im zum Verständnis notwendigen Kontext darstellen. Nur so können die Datennutzenden nachvollziehen, wie das Ergebnis zu verstehen ist. Um diese Einordnung für die Datennutzenden vornehmen zu können, müssen die Datenverarbeitenden dazu in der Lage sein nachzuvollziehen, wie das Ergebnis zustande gekommen ist bzw. welcher Input zu welchem Output führt und welche Trainingsdaten verwendet wurden (siehe Tabelle 4).

Diese administrativen Anforderungen sorgen dafür, dass die Datennutzenden in Bezug auf den Umgang mit den Daten, denen sie im Datenökosystem ausgeliefert sind, selbstbestimmt handeln können und sie somit nicht als bloßes Mittel gebraucht werden. Da die Auferlegung eines Risikos nur dann zulässig ist, wenn niemand als bloßes Mittel gebraucht wird, ist die Erfüllung der administrativen Anforderungen eine notwendige Bedingung für die Zulässigkeit der Risikoauferlegung.

992 Hansson 2003:305
993 Kant 2008:65 [429]
994 Mittelstadt et al. 2016:6

Tabelle 4 Administrative Anforderungen für die Umsetzung von Transparenz

Transparenz Definition	**Transparenz Ziele**	**Umsetzung des Ziels**	**Administrative Anforderungen**
Transparenz = Zugänglichkeit und Verständlichkeit[995]	Algorithmische Ergebnisse sollen *verständlich* sein.	Nur durch die *Erklärung der Entstehung* des algorithmischen Ergebnisses sind wir in der Lage, seine Bedeutung zu verstehen.	*Entstehung des Ergebnisses verstehen:* Welcher Input führt zu welchem Output, und welche Trainingsdaten wurden verwendet?
	Algorithmische Ergebnisse sollen *zugänglich* sein.	Algorithmische Ergebnisse werden den Datennutzenden zur Verfügung gestellt, *samt leicht verständlicher Erklärung.*	Algorithmische Ergebnisse im *zum Verständnis notwendigen Kontext* darstellen.

Aus der Pflicht zur Transparenz der algorithmischen Ergebnisse und den zur Umsetzung notwendigen administrativen Anforderungen lässt sich ein klarer Vorteil erkennen für den Einsatz von Wenn-Dann-Algorithmen. Durch die Klarheit der vorgegebenen Entscheidungsregeln sind die Ergebnisse transparent und können problemlos zugänglich gemacht und verständlich dargestellt werden.

Es lassen sich aber nicht alle Sachverhalte auf klare Regeln herunterbrechen. Oft gibt es einen Ermessensspielraum, sodass für eine Eingabe mehrere Ausgaben möglich sind. Zum Beispiel hat eine Behörde im Rahmen von vorgegebenen Bauvorschriften und Umweltstandards den Ermessensspielraum, eine Baugenehmigung zu erteilen oder nicht. Oder wenn ein Strafmaß nicht eindeutig festgelegt ist, hat ein Richter, innerhalb des Gesetzesrahmens, den Ermessensspielraum, die Höhe des Strafmaßes zu bestimmen. Eine andere Situation, in der die Ausgabe nicht vorgegeben werden kann, ist die Suche nach von Menschen unentdeckten Mustern in Datensätzen. Eine solche Mustersuche kann zum Beispiel angewendet werden bei der Einstellung von Personal. So können zum Beispiel, wie bei

995 Mittelstadt et al. 2016:6

Amazon im Jahr 2014, Muster identifiziert werden in Bewerbungen, die in der Vergangenheit erfolgreich waren. Auf diese Weise werden Merkmale des Erfolgs abgeleitet und zu Entscheidungskriterien für neue Bewerbungen gemacht. Nun können aus vorliegenden Bewerbungen die Besten herausgefiltert werden, ohne vorzugeben was eine Bewerbung gut macht.[996]

Sobald es keine klare Wenn-Dann Beziehung gibt, sondern ein Ermessensspielraum gilt oder eine Mustersuche das Ziel ist, kann ein regulärer Wenn-Dann-Algorithmus nicht mehr eingesetzt werden. Um in einem solchen Fall trotzdem ein Ergebnis zu erhalten, ist der Einsatz von KI-Systemen möglich. Expert:innen wie Datenverarbeitende kennen die unter 7.2 beschriebenen Praktiken, mit deren Hilfe sie die Glaubwürdigkeit von Algorithmen überprüfen können. Datennutzende, welche die Ergebnisse des Modells nutzen oder von ihnen betroffen sind, sind hierbei jedoch nicht einbezogen, sodass es für die breite Öffentlichkeit schwer ist, Algorithmen zu überprüfen. Dadurch bleiben epistemische Asymmetrien zwischen Datenverarbeitenden und Datennutzenden meist bestehen.[997]

Das ist deshalb problematisch, weil Datennutzende epistemisch abhängig sind von KI-Systemen, wenn sie sich auf ihre Ergebnisse verlassen (müssen). Auch von anderen Technologien, wie zum Beispiel Thermometern, sind wir epistemisch abhängig. Aber die Ergebnisse von Thermometern sind deskriptiv und erfordern offensichtlich Kontextualisierung. Algorithmische Ergebnisse hingegen fordern oft direkt zum Handeln auf – zum Ablehnen eines Wohngeldantrages oder zum Verfolgen eines bestimmten Berufsweges. Jedoch treffen in der Regel nicht die Algorithmen die endgültige Entscheidung, da Menschen beteiligt sind, die ihre Entscheidungen auf der Grundlage der algorithmischen Ergebnisse treffen. Die menschlichen Entscheidungstragenden sind laut Grote et al. nicht nur passive Empfänger von Informationen, sondern Expert:innen in eigener Sache.[998] Im Kontext eines Datenökosystems läge die endgültige Entscheidung über die Ablehnung eines Wohngeldantrages demnach bei den Mitarbeitenden der Wohngeldbehörde.

996 Stahl et al. 2023:10f.
997 Kolkman 2022:104
998 Grote et al. 2024:5

Allerdings ist es, wie unter 7.1 erläutert, laut van den Hoven oft nicht möglich, anders zu entscheiden, als es der Algorithmus empfiehlt: „Users may become epistemically dependent upon the system they are working with, they cannot but cast their accounts of what they did and why they did it wholly in terms of the system. They may me unable to put forward 'system independent reasons'."[999] Aktuell sind Verwaltungsmitarbeitende Expert:innen für die Frage, wann ein Wohngeldantrag abgelehnt wird. Sie kennen die Gesetzeslage und arbeiten sich durch die Antragsunterlagen. Doch wenn ein Algorithmus eingesetzt wird, um diese Arbeit zu erleichtern, dann ist es nicht mehr notwendig, dass die Mitarbeitenden die Antragsunterlagen im Detail betrachten – es kann nicht einmal gewollt sein, denn sonst wäre der Einsatz des Algorithmus ja nicht notwendig oder sinnvoll. Dadurch entsteht zusätzlich die Gefahr des Verlustes dieser Expert:innenkenntnisse (deskilling), da sie nicht mehr regelmäßig gebraucht werden.[1000] So kann es über die Zeit sogar zur Herausforderung werden, die algorithmischen Ergebnisse im Zweifel händisch zu überprüfen.

Beim Einsatz von KI-Systemen muss also bedacht werden, dass es zusätzliche Arbeit erfordert, um die Pflicht zur Transparenz erfüllen zu können. Deshalb empfiehlt es sich, für alle Anwendungsfelder ohne Entscheidungsspielraum ausschließlich Wenn-Dann-Algorithmen zu verwenden. So kann zum Beispiel die Ablehnung eines Wohngeldantrages mit der Gesetzeslage begründet werden, welche dem Algorithmus als Entscheidungskriterium vorgegeben wurde. KI-Systeme hingegen sollten nur dann eingesetzt werden, wenn ihr Vorgehen bei der Beantwortung der Frage hilfreich ist, wenn der zugrundeliegende Datensatz wirklich zu hilfreichen Aussagen führen kann und wenn sich im Nachhinein erklären lässt, welcher Input zu welchem Output geführt hat.

999 Van den Hoven 1998:104
1000 Rafner et al. 2021

8 Verantwortung

Um herausfinden zu können, wer bei der Gestaltung eines Datenökosystems die Verantwortung für die Risikoauferlegung, sowie den Schutz der Privatsphäre von Datenquellen und die Bereitstellung der Transparenz für Datennutzende trägt oder tragen sollte, wird im Folgenden zuerst geklärt, was Verantwortung bedeutet. Mit Hilfe dieser Definition soll dann erörtert werden, wer im Datenökosystem Verantwortung trägt und somit die Risiken bezüglich Privatsphäre und Transparenz, sowie die Zulässigkeit der Risikoauferlegung bei der Gestaltung des Datenökosystems beachten muss. Die Begründung der Notwendigkeit eines Konzepts der Verantwortung für die ethische Gestaltung eines Datenökosystems ist, dass das Fehlen von verantwortungsvollem Handeln zu Schaden führen kann und dass es im Fall eines Schadens wichtig ist, dass jemand dafür zur Rechenschaft gezogen werden kann.

Denn Menschen haben, nach Ansicht des Technologiephilosophen John Danaher, ein Interesse daran, dass jemand im Schadensfall die Verantwortung übernimmt – wenn sie geschädigt werden, suchen sie nach einem schuldigen Täter, der eine Strafe verdient hat.[1001] Auch der Technologiephilosoph und -ethiker Maximilian Kiener identifiziert dieses Interesse an der Verantwortungsübernahme: „When we suffer harm through other people's action, we have an interest that *some* people not only plan to or *de facto* explain to us what happened, follow up on us, etc., but also that they *morally owe* this to us."[1002] Wer für seine Handlung Verantwortung übernimmt und dadurch anerkennt, dass ein bestimmter Schaden eine zwischenmenschliche moralische Anerkennung erfordert, ermöglicht den Geschädigten ein aktives Mitspracherecht, da sie entscheiden können, ob sie die Verantwortlichen von ihrer moralischen Verpflichtung entbinden. Wenn aber für entstandenen Schaden kei-

1001 Danaher 2016:299
1002 Kiener 2022:584

ne Verantwortung übernommen wird, wird dieser als natürliches Unglück behandelt, und die Geschädigten werden ohne Kontrolle über die moralischen Verpflichtungen der Verantwortlichen entwürdigt und objektiviert.[1003] Dem KI-Ethiker Sven Nyholm zufolge kann die Möglichkeit, dass wir nicht in der Lage sind, jemanden zu finden, den wir für etwas verantwortlich machen können, selbst als negatives Ergebnis angesehen werden.[1004]

Die KI-Ethikerin Ann-Katrien Oimann und der Technologiephilosoph Fabio Tollon halten es zudem für das öffentliche Vertrauen in Demokratien für unerlässlich, dass insbesondere bei negativen Folgen angemessen Verantwortung zugewiesen werden kann.[1005] Zudem ist es der Politologin Iris Marion Young zufolge in einem System der moralischen Regeln und der rechtlichen Verantwortlichkeit wichtig, dass alle Akteure wissen, dass sie als einzelne Akteure angeklagt oder zur Verantwortung gezogen werden können.[1006] Auf diese Art führt Verantwortung zu bedachtem Verhalten auf Seiten der Handelnden, und sie sorgt für ein Gefühl der Sicherheit auf Seiten der Behandelten. Wenn zum Beispiel der Hersteller eines Haushaltsgeräts Verantwortung für sein Produkt übernimmt, dann ist es wahrscheinlicher, dass Kund:innen an die Funktionsfähigkeit und Sicherheit des Produktes glauben, als wenn der Hersteller jegliche Verantwortung für mögliche negative Folgen seines Produktes von sich weist. Auch im Datenökosystem geht es bei der Übernahme von Verantwortung darum, dass Datenquellen und Datennutzende sicher sein können, dass sich jemand um den Schutz ihrer Privatsphäre beziehungsweise um die Ermöglichung von Transparenz kümmert und die Zulässigkeit der Risikoauferlegung im Blick hat.

8.1 Definition

Verantwortung heißt im Lateinischen „responsibilitas". Zur Herleitung des Begriffs schreiben die Bildungsphilosophen Holger Burckhart und Jürgen Nielsen-Sikora: „Das lateinische Verb *respondere*

1003 Kiener 2022:584
1004 Nyholm 2024:192
1005 Oimann, Tollon 2024:5
1006 Young 2011:105f.

bedeutet „antworten“ und meint vor allem die Verpflichtung, Antworten auf die von einem Gericht gestellten Fragen zu geben.“[1007] Auch im Deutschen enthält der Begriff Verantwortung das Wort „antworten“. Im juristischen Sprachgebrauch bedeutet Antworten, vor einem Gericht Rechenschaft abzulegen.[1008] Das heißt, es wird gerechtfertigt, aus welchem Grund Handlungen ausgeführt wurden oder nicht. Sich Verantworten heißt also wörtlich, dass man sein Handeln im Nachhinein rechtfertigt.

Im Folgenden soll es nicht um rechtliche Verantwortung gehen – also um die Pflicht sich an juridische Rechte zu halten – und es soll auch nicht um rein kausale Verantwortung gehen, welche auch Gegenständen oder Naturereignissen zugeschrieben werden kann. Es soll um moralische Verantwortung gehen, die nur durch moralische Akteure getragen werden kann und für die es kein (rechtliches) Durchsetzungssystem gibt.[1009] Moralische Verantwortung tragen heißt dementsprechend, dass man dazu in der Lage sein sollte, auf Fragen über die zu verantwortende Handlung zu antworten und somit für die Handlung sowie ihre Folgen Rechenschaft abzulegen. Man ist also vor Anderen oder sich selbst für das eigene Handeln und seine Folgen verantwortlich.[1010]

Es wäre jedoch überfordernd, wenn diejenigen, die Verantwortung tragen, für alle möglichen Folgen ihres Handelns (und Nicht-Handelns) Rechenschaft ablegen müssten. Auf Grund von klaren Kausalketten wären zum Beispiel diejenigen, die den Verbrennermotor und das Auto für den Individualverkehr erfunden und produziert haben, mit verantwortlich für den Klimawandel – dabei waren diese drastischen Folgen damals nicht absehbar. Da viele zukünftige Folgen, insbesondere im Bereich der Technologieentwicklung, heute kaum oder noch gar nicht absehbar sind, muss der Verantwortungsbegriff geschärft werden.

Die Gesinnungsethik besagt dem Soziologen Max Weber zufolge, dass es sich anhand der inneren Einstellung gegenüber einer Handlung entscheidet, ob jemand Verantwortung für sein Handeln tragen muss. Eine reine Gesinnung haben bedeutet, dass man das Richtige

1007 Burckhart, Nielsen-Sikora 2020:177
1008 Ebd.
1009 Isaacs 2016:4
1010 Betzler, Scherrer 2016:2; Burckhart, Nielsen-Sikora 2020:183

tun will und eine Handlung ausführt, weil man davon überzeugt ist: „Wenn die Folgen einer aus reiner Gesinnung fließenden Handlung üble sind, so gilt ihm nicht der Handelnde, sondern die Welt dafür verantwortlich (…)."[1011] Das bedeutet, solange man mit besten Absichten das Richtige tun wollte, trägt man keine Verantwortung für negative Folgen der Handlung. Nun treten aber auch bei Handlungen aus reiner Gesinnung negative Folgen auf. So wurden zum Beispiel die Daten psychotherapeutischer Patient:innen in Finnland womöglich deshalb digitalisiert gespeichert, weil das zur Übersichtlichkeit bei der Behandlung beitragen sollte – man war davon überzeugt, das Richtige zu tun. Trotzdem wurden die Daten gestohlen und die Patient:innen erpresst.[1012] Ich halte es für problematisch, dass diejenigen, die die Daten digitalisiert gespeichert haben, keine Verantwortung für die negativen Folgen tragen sollen, weil sie aus reiner Gesinnung gehandelt haben.

Weber würde mir zustimmen und hat deshalb die Verantwortungsethik entwickelt. Diese besagt „(…) daß man für die (voraussehbaren) Folgen seines Handelns aufzukommen hat."[1013] Das heißt, für alle Folgen meines Handelns, mit denen ich im Vorhinein rechnen konnte, muss ich Verantwortung tragen. Auf diese Weise werden die absehbaren Folgen meines Handelns zum Kriterium ebendieser Handlungen. Menschen rufen oft, trotz guter Absichten, mit ihrem Handeln negative Folgen hervor. Wenn die Folgen zum Zeitpunkt der Handlung absehbar waren, sollte der handelnde Mensch gemäß der Verantwortungsethik den Fehler bei sich suchen und Verantwortung übernehmen, anstatt auf die eigene reine Gesinnung zu verweisen. Laut Weber macht aber erst die Kombination aus reiner Gesinnung und der Beachtung von absehbaren Folgen eine Handlungsentscheidung verantwortungsbewusst: „Insofern sind Gesinnungsethik und Verantwortungsethik nicht absolute Gegensätze, sondern Ergänzungen, die zusammen erst den echten Menschen ausmachen (…)."[1014] Wer also mit reiner Gesinnung und alle absehbaren Folgen bedenkend handelt, der sollte im Fall des Eintretens negativer Folgen dafür Rechenschaft ablegen.

1011 Weber 1926:58
1012 Looi et al. 2025:107f.; Ghanbari, Koskinen 2024:1f.
1013 Weber 1926:58
1014 Ebd.:66

Dabei ist es den Technologiephilosophen Vincent Blok und Pieter Lemmens zufolge aber wichtig, dass die Handlungsfolgen, für die Verantwortung getragen werden soll, voraussehbar waren: „If there is practically no way to predict or foresee such consequences, however, all talk about responsibility in this context would seem groundless and misleading."[1015] Niemand kann verantwortlich gemacht werden für Folgen, die er oder sie nicht voraussehen konnte. Van den Hoven bestätigt, dass Menschen nicht immer unter vollständigem Wissen handeln: „Sometimes we have to act and choose under conditions of uncertainty or ignorance and take responsibility for what we do relative to what we know, and be held responsible relative to what we knew and what we could have known."[1016] Was also von Natur aus epistemisch unzugänglich ist, kann nicht gewusst oder vorausgesehen werden und dafür kann man dementsprechend auch nicht zur Verantwortung gezogen werden. Aber man ist sehr wohl verantwortlich, wenn man sich vor Wissen verschließt, um die Folgen nicht verantworten zu müssen oder gar nicht erst über die Grenzen des eigenen Wissens nachdenkt. Unwissenheit ist dann schuldhaft, wenn man diese hätte erkennen und verhindern können. Wenn man sich aber nicht ausreichend über mögliche Folgen der eigenen Handlungen erkundigt, obwohl das möglich wäre, dann verbessert man absichtlich seine epistemische Situation nicht und kann für Folgen, die absehbar gewesen wären, zur Verantwortung gezogen werden.[1017]

Neben der Berücksichtigung von voraussehbaren Folgen ist den praktischen Philosophinnen Monika Betzler und Nina Scherrer zufolge eine weitere Bedingung für das Tragen von Verantwortung, dass die betreffende Person Kontrolle über die Handlung hat: „(...) unter Zwang oder in Unwissenheit kontrollieren Personen ihr eigenes Handeln nicht und sind entsprechend nicht verantwortlich."[1018] In der Literatur ist jedoch umstritten, was es bedeutet, Kontrolle über seine Handlung zu haben – die Betzler und Scherrer fassen die verschiedenen Ansätze sehr übersichtlich zusammen.[1019] Wäh-

1015 Blok, Lemmens 2015:26
1016 Van den Hoven 2013:80
1017 Betzler, Scherrer 2016:11
1018 Ebd.:2
1019 Ebd.

rend die einen meinen, man müsse volitionale Kontrolle über seine Handlung haben und diese willentlich verursachen[1020], meinen andere, dass man Führungskontrolle über die eigene Handlung haben sollte und damit hypothetisch die Möglichkeit, anders zu handeln, wenn es Gründe dafür gibt[1021]. Ein weiterer Ansatz besagt, dass man erst dann Kontrolle über die eigene Handlung hat, wenn man tatsächlich alternative Handlungsmöglichkeiten hat und selbst evaluiert, aus welchem Grund man welche Handlungsmöglichkeit auswählt.[1022] Zudem gibt es den Ansatz der rationalen Beziehungen mit der Idee, dass man dann verantwortlich ist, wenn man vernünftigerweise nach einer Rechtfertigung gefragt werden kann.[1023]

Es gibt also sehr unterschiedliche Vorstellungen davon, welchen Grad an Kontrolle man über eine Handlung haben muss, um für sie verantwortlich sein zu können. Es herrscht jedoch Einigkeit darüber, dass ein Akteur dann die Verantwortung für seine Handlung und deren absehbare Folgen trägt, wenn er als ihr Urheber angesehen wird.[1024] Die Technologiephilosophen Johannes Himmelreich und Sebastian Köhler sind aber zudem davon überzeugt, dass es ausreicht, wenn eine Beziehung den betreffenden Akteur mit der Handlung und ihren absehbaren Folgen verbindet, um verantwortlich zu sein: „But, this relation need not be a strong conception of *control* or even a causal-like relation."[1025] Denn man könne auch verantwortlich sein für die Handlung einer anderen Person, zum Beispiel des eigenen Kindes, oder müsse eventuell Schadensersatz zahlen für Folgen einer Handlung, über die man keine Kontrolle hatte, zum Beispiel als Vorsitzende:r eines Unternehmens.[1026] Himmelreich und Köhler verstehen die Beziehung zwischen verantwortlicher Person und zu verantwortender Handlung daher mit Hilfe einer Konzeption von Nyholm eher als Beaufsichtigung (supervision).[1027] In dem Bereich, in dem man Aufsicht führt, müsse man für die entsprechenden Handlungen und deren absehbare Folgen einstehen und diese

1020 Frankfurt 1971
1021 Fischer, Ravizza 1998
1022 Kane 1996
1023 Smith 2005
1024 Mieth, Bambauer 2016:4
1025 Himmelreich, Köhler 2022:24
1026 Ebd.:24f.
1027 Nyholm 2018:1216; Dazu mehr in Unterkapitel 8.1.2.

Kausalbeziehung ließe sich auf die Bedingungen für Verantwortung übertragen.[1028]

Auch dem Rechtsphilosophen Herbert Lionel Adolphus Hart zufolge muss eine Person nicht die direkte Ursache einer Handlung sein, um für diese verantwortlich zu sein. Denn die Verbindung mit der Handlung kann auch ohne direkte kausale Verbindung ausreichen, um eine Person verantwortlich oder rechtlich haftbar zu machen: „(...) it is common in all systems of civil law for men to be made to pay compensation for injuries caused by others, generally their (...) employees."[1029] Zum Beispiel wenn Angestellte Schaden verursachen oder jemand durch den Einsatz von fehlerhaften Maschinen zu Schaden kommt, können die Arbeitgebenden dafür haftbar und dementsprechend im rechtlichen Sinne verantwortlich sein.[1030] Hart hat im Jahr 1968 mit den fehlerhaften Maschinen sicherlich noch keine Algorithmen gemeint, aber das Beispiel lässt sich sehr gut auf diese übertragen. Zudem tragen auch Eltern Verantwortung für ihre Kinder und Tierbesitzer tragen Verantwortung für ihre Tiere – weil sie eine bestimmte Rolle in einer sozialen Organisation einnehmen oder diese ihnen zugeschrieben wurde.

Mieth und Bambauer widersprechen jedoch der Abgrenzung von kausaler Verantwortung, weil die Haftung von Arbeitgebern, Eltern und Tierbesitzern eben doch auf die Möglichkeit des Einflusses zurückzuführen ist: „Dennoch basieren auch diese Formen der Haftungsverantwortung indirekt auf der Zuschreibung von kausaler Verantwortung, denn Eltern haften für ihre Kinder und Tierbesitzer für ihre Tiere nur deswegen, weil ihnen eine weitreichende kausale Einflussnahme auf das Verhalten ihrer Kinder bzw. Tiere unterstellt wird."[1031] In gleicher Weise können auch Arbeitgebende auf ihre Angestellten und auf die von ihnen eingesetzten Maschinen Einfluss nehmen. Insofern haben Menschen in bestimmten Rollen eine gewisse Kontrolle über die Handlungen der Kinder und der Angestellten, sowie das Verhalten der Tiere und die Ergebnisse der Maschinen. Diese Kontrolle ermöglicht die Zuschreibung von Verantwortung.

1028 Himmelreich, Köhler 2022:24
1029 Hart 1968:216
1030 Ebd.:216, 222
1031 Mieth, Bambauer 2016:8

Dem analytischen Philosophen Peter Strawson zufolge sind wir jedoch nicht für eine Handlung verantwortlich auf Grund unserer Kontrolle über diese Handlung, sondern in seinen Augen sind es unsere Praktiken, uns gegenseitig moralische Verantwortung zuzuschreiben, die festlegen, ob jemand moralisch verantwortlich ist. Somit werden „reaktive Einstellungen" (reactive attitudes) bezüglich einer Handlung, wie zum Beispiel Ärger, zur wichtigsten Methode, mit der wir den Akteuren in angemessener Weise moralische Verantwortung zuschreiben können.[1032] Es gäbe also einen Unterschied zwischen verantwortlich sein (being responsible) und zur Verantwortung gezogen werden (holding responsible). Verantwortlich sein sei unabhängig von „reaktiven Einstellungen" der Menschen (response independent), sodass eine Person verantwortlich für eine Handlung sei, wenn sie Kontrolle über diese Handlung hat und deren Folgen absehen konnte. Nach diesem Ansatz können zum Beispiel Kleinkinder nicht als moralisch verantwortlich angesehen werden, weil ihnen die entsprechenden Fähigkeiten für diese Zuschreibung fehlen. Zur Verantwortung gezogen werden sei jedoch abhängig von „reaktiven Einstellungen" anderer Menschen (response dependent), sodass eine Person als verantwortlich für eine Handlung gelte, wenn die Menschen sie dafür zur Verantwortung ziehen. Dabei solle die reaktive Einstellung angemessen sein, in dem Sinne, dass es einen bestimmten Grund gibt, diese Einstellung zu beziehen. Nach diesem Ansatz ist es also möglich, empirische Daten heranzuziehen, um die Regeln für die Anwendung des Konzepts der Verantwortung zu bestimmen.[1033] Dieser Ansatz von Strawson weist laut Oimann und Tollon daraufhin, dass das Konzept der moralischen Verantwortung davon geprägt ist, dass wir soziale Menschen sind, die in einer bestimmten moralischen Gemeinschaft leben.[1034]

Ich argumentiere jedoch, dass solche reaktiven Einstellungen als Methode zur Zuschreibung von Verantwortung immer nur dann angemessen sind, wenn ihr Grund ist, dass eine Verbindung zwischen dem betreffenden Menschen und der Handlung, inklusive ihrer absehbaren Folgen, besteht. Das hieße dann, eine Person könnte auch nach Strawsons Ansatz nur dann für eine Handlung zur Verantwor-

1032 Strawson 1962; De Mesel, Heyndels 2019; Oimann, Tollon 2024:3
1033 Oimann, Tollon 2024:3f.
1034 Ebd.:12

tung gezogen werden, wenn sie einen Einfluss auf die Handlung und damit eine gewisse Kontrolle über die Handlung hatte. Denn wenn die Person zu der Handlung gar keine Verbindung hatte, dann findet sich kein Grund für eine reaktive Einstellung, wie zum Beispiel Ärger, der Person gegenüber. Daraus schlussfolgere ich, dass eine Person nur dann für eine Handlung zur Verantwortung gezogen werden kann, wenn die Person in Beziehung zu der Handlung steht, indem sie, mindestens in einer beaufsichtigenden Rolle, Kontrolle über die Handlung hat.

8.1.1 Schuld und Strafe

Auf der Grundlage dieser Relation ist es laut Betzler und Scherrer angemessen, die Person für die Handlung und ihre absehbaren Folgen zu loben oder zu tadeln.[1035] Der Gedanke findet sich auch bei Burckhart und Nielsen-Sikora: „Denn mache ich jemanden für eine bereits vergangene oder aber erst noch zu treffende Entscheidung verantwortlich, geht damit die Erwartung einer möglichen Schuldübernahme einher."[1036] Jemandem die Schuld zu geben bedeutet laut Menges, auf eine bestimmte Art und Weise wütend auf die Person zu sein. In gleichem Maße bedeute sich selbst die Schuld zu geben, auf diese bestimmte Art und Weise wütend auf sich selbst zu sein oder sich schuldig zu fühlen.[1037] Menges argumentiert, dass verantwortlich sein für eine verwerfliche Handlung nicht nur meint, dass man deshalb ein angemessenes Ziel von Schuldzuweisungen ist, sondern diese beiden Fakten seien gleichzusetzen. Sein sogenannter „No-Difference"-Ansatz besagt also, dass wenn eine Handlung, für die man verantwortlich ist, verwerflich ist, man folglich das angemessene Ziel einer Schuldzuweisung ist.[1038] In gleicher Weise gilt, dass wenn man das angemessene Ziel einer Schuldzuweisung für

1035 Betzler, Scherrer 2016:2
1036 Burckhart, Nielsen-Sikora 2020:178
1037 Menges 2020b:394
1038 Ebd.:396

eine verwerfliche Handlung ist, ist man gleichzeitig verantwortlich für diese Handlung.[1039]

Die Philosophin und Frauenrechtlerin Edith Stein ist der Meinung, dass durch freie Handlungen eine Schuld entstehen kann, welche nach Strafe verlangt: „Verantwortlichkeit bezeichnet die Fähigkeit, Urheber einer Schuld und damit Objekt der ihr gebührenden Strafe zu sein."[1040] An dieser Stelle wird erneut deutlich, dass die Kontrolle über eine Handlung und deren absehbare Folgen notwendig ist, um für diese verantwortlich zu sein, da es ungerecht oder unverdient wäre, eine Person für eine Handlung zu tadeln oder zu bestrafen, wenn sie keine Kontrolle darüber hätte.[1041] Auch in europäischer Rechtsprechung bedeutet verantwortlich zu sein strafbar zu sein im Falle eines Verstoßes. Wenn über eine Geldbuße und ihre Höhe entschieden wird, kann unter anderem der Grad an Verantwortung des Akteurs berücksichtigt werden.[1042]

Die Verantwortung für negative Folgen zieht jedoch nicht immer eine Strafe nach sich. Unterschiedliche Handlungsfolgen erfordern unterschiedliche Reaktionen seitens der Akteure, deren Handeln diese Folgen hervorruft.[1043] Laut Kiener entsteht durch Verantwortung als die Pflicht sich zu erklären, auch die Pflicht sich zu entschuldigen, sich um das Wohlergehen derjenigen zu kümmern, die geschädigt wurden, sowie Vorsichtsmaßnahmen zu ergreifen, damit sich ähnlicher Schaden nicht wiederholt.[1044] So eine Entschuldigung könne aber nur dann angemessen sein, wenn man wirklich verant-

1039 Iris Marion Young würde dem ‚No-Difference'-Ansatz widersprechen, da sie der Meinung ist, dass jemand gleichzeitig verantwortlich und nicht schuldig sein kann (Young 2011:105–111). Die Ursache für die Uneinigkeit liegt in der Perspektive, denn der ‚No-Difference'-Ansatz betrachtet die Vergangenheit und besagt, dass wer für schon entstandenen Schaden verantwortlich ist, immer auch als schuldig bezeichnet werden kann und andersherum. Young hingegen betrachtet die Zukunft und ihr Ansatz besagt, dass Menschen dafür verantwortlich sein können, die Zukunft besser zu gestalten, auch wenn sie nicht als schuldig bezeichnet werden können für den aktuellen Zustand. Der Widerspruch entsteht also durch eine unterschiedliche Definition von Verantwortung, welche im Folgenden noch eingehend thematisiert wird.

1040 Stein 2014:31

1041 Menges 2020c:396

1042 Art. 99, Abs. 7 (g) EU AI Act; Art. 100, Abs. 1 (b) EU AI Act

1043 Wolf 2001:16

1044 Kiener 2022:582

wortlich ist und nicht nur positive Konsequenzen dadurch erzeugen will – ohne Verantwortung handele es sich nur um einen Ausdruck von Bedauern.[1045] Laut dem Moralphilosophen und Utilitaristen Henry Sidgwick ist es jedoch richtig, eine Person für eine Handlung mit negativen Folgen nicht mit einer Entschuldigung davon kommen zu lassen, sondern sie zu bestrafen, damit die Angst vor Strafe sie und andere davon abhält, in Zukunft wieder so zu handeln.[1046]

Laut Kiener kann ein Mensch allerdings nur dann rechtlich bestraft werden, wenn er einen Fehler begangen hat und deshalb eine Schuldzuweisung berechtigt ist: „(...) punishment is also connected to fault."[1047] Zur Unterstützung dieser These zieht er unter anderem die Moralphilosophin und Sozialepistemologin Miranda Fricker sowie den politischen Philosophen Gary Watson heran. Denn auch Fricker ist der Meinung, dass Schuldzuweisungen unangebracht sind, wenn man ohne eigenes Verschulden negative Folgen verursacht. Wenn zum Beispiel ein LKW-Fahrer sehr vorsichtig fährt, aber durch ein tragisches Unglück ein Kind überfährt, ist das laut Fricker nicht der Fehler des Fahrers und deshalb könne ihm auch keine Schuld zugewiesen werden: „If no fault, then no appropriate blame."[1048] Auch laut Watson bedeutet jemandem moralische Schuld zuzuweisen, dass diese Schuld auf einen moralischen Fehler des Handelnden zurückzuführen ist.[1049] Kiener schließt daraus, dass man einer Person nicht die Schuld geben kann, wenn diese Person keinen Fehler gemacht hat. Es mache nicht einmal einen Unterschied, wenn die Person erklärt, dass sie die Schuld akzeptiert: „Hence, if there is to be genuine blame, there must be fault too."[1050] Ein Fehler, der einen schuldfähig macht, entstehe dabei durch ungerechtfertigtes oder nicht zu entschuldigendes Handeln. Solche Fehler seien das, was moralische Schuld von Enttäuschung unterscheide.[1051]

1045 Kiener 2022:590
1046 Sidgwick 2011 [1874]:50, 57
1047 Kiener 2022:580
1048 Fricker 2016:170
1049 Watson 2004:266
1050 Kiener 2022:578f.
1051 Ebd.

8.1.2 Verantwortungslücke

Wenn also durch eine Handlung ein Schaden entsteht, dieser Schaden aber nicht auf einen Fehler in Form von ungerechtfertigtem oder nicht zu entschuldigendem Handeln zurückzuführen ist, dann entsteht laut Kiener eine Verantwortungslücke (responsibility gap).[1052] Laut Oimann und Tollon war der Technologiephilosoph Andreas Matthias im Jahr 2004 der Erste, der von einer solchen Verantwortungslücke geschrieben hat.[1053] Seiner Meinung nach besteht genau dann eine Verantwortungslücke, wenn niemand genügend Kontrolle über die Handlungen der Maschine hat, um die Verantwortung dafür übernehmen zu können.[1054] Die Idee der Verantwortungslücken ist aber nicht auf den Kontext des Einsatzes von Maschinen (wie zum Beispiel algorithmische Systeme) beschränkt. Unabhängig vom Maschinenkontext gibt es auch dann keine hauptverantwortliche Person, wenn an einer Entscheidung zu viele Menschen beteiligt sind, von denen jeder einen kleinen Teil der Verantwortung trägt. Solche Fälle werden auch als das „Problem der vielen Hände" (problem of many hands) bezeichnet.[1055]

Dem politischen Philosophen sowie Bio- und Technologieethiker Robert Sparrow zufolge entsteht eine Verantwortungslücke, wenn negative Folgen einer Handlung den Wunsch erzeugen, dass jemand zur Verantwortung gezogen werden kann, aber niemand die Bedingung der Kontrolle über die Handlung und ihre absehbaren Folgen erfüllt.[1056] Denn solange eine Person die verwerflichen Folgen nicht vorhersehen oder abwenden konnte, kann sie dafür nicht angemessen verantwortlich gemacht werden.[1057] Diese Situation tritt zum Beispiel dann auf, wenn KI-Systeme zum Einsatz kommen, deren Ergebnisse sich menschlicher Kontrolle entziehen. Denn die Verbindung zwischen den Programmierenden und den Ergebnissen des Systems, welche die Zuschreibung von Verantwortung begründen würde, wird durch die Fähigkeit des Systems zu lernen unter-

1052 Kiener 2022:579
1053 Oimann, Tollon 2024:10; Matthias 2004
1054 Matthias 2004:177
1055 van de Poel et al. 2015; Nyholm 2024:194
1056 Sparrow 2007:70; Oimann, Tollon 2024:5
1057 Matthias 2004:177

brochen.[1058] Aber auch die algorithmischen Systeme, die in einem kausalen Verhältnis zu den Ergebnissen stehen, können keine Verantwortung tragen.[1059] Denn algorithmische Systeme haben weder volitionale Kontrolle über ihre Ergebnisse[1060], noch haben sie hypothetisch oder tatsächlich die Möglichkeit, andere Ergebnisse zu produzieren oder können Gründe angeben für ihr Vorgehen[1061]. Außerdem können algorithmische Systeme nicht vernünftigerweise nach einer Rechtfertigung gefragt werden[1062], da diese keine Diskussion über die Gründe für oder gegen ihre Ergebnisse führen können.[1063]

Gemäß Sparrow sind Verantwortungslücken bei dem Einsatz von KI-Systemen unvermeidbar und problematisch.[1064] Im Gegensatz dazu gibt es auch Autor:innen, die Verantwortungslücken willkommen heißen. So kann es laut den Digitalethikern Lauritz Munch und Jakob Mainz sowie dem Technologiephilosophen Jens Christian Bjerring wünschenswert sein, Entscheidungsprozesse zu automatisieren und menschliche Eingriffe aus kritischen Entscheidungsprozessen zu entfernen, weil dann kein Mensch verantwortlich für potenzielles Fehlverhalten ist. Dies sei deshalb wünschenswert, weil es schlecht sein kann, für Fehlverhalten verantwortlich zu sein. Denn Verantwortung für Fehlverhalten bringe Dinge mit sich wie Schuldzuweisung, ein schlechtes Gewissen oder sogar Strafen. Laut Munch et al. sind insbesondere Strafen in der Regel schlecht für die Person, die für Fehlverhalten verantwortlich gemacht wird: „Provided that we care about directing human agency away from wrongdoing, we thus have a *pro tanto* reason to prefer AI decision-makers over human decision-makers (…)."[1065] Munch et al. sind der Meinung, dass durch den Einsatz von algorithmischen Systemen Menschen die Entscheidungslast genommen werden kann und sie dadurch Fehlverhalten vermeiden können. Auch Danaher ist der Meinung, dass es hilfreich sein kann, Entscheidungen an Maschinen zu delegieren, insbesondere wenn es sich um moralische Dilemmata handelt, bei denen es unmöglich ist,

1058 Sparrow 2007:70
1059 Oimann, Tollon 2024:5
1060 Frankfurt 1971
1061 Fischer, Ravizza 1998; Kane 1996
1062 Smith 2005
1063 Nyholm 2018:1209
1064 Sparrow 2007:74
1065 Munch et al. 2023:6

alle Erwägungen perfekt auszubalancieren. Mit solchen Entscheidungen sei eine Last der Verantwortung verbunden, welche durch die Verantwortungslücke gemindert werden könne.[1066]

Eine Entscheidung durch Algorithmen anstatt durch Menschen verhindert vielleicht ein schlechtes Gewissen oder unangenehme Strafen, aber es verhindert nicht ein Fehlverhalten und seine negativen Folgen. Das sieht auch Kiener kritisch: „(…) there is a human-induced harm without moral answerability, which leaves our answerability interests frustrated."[1067] Wenn also eine Person von algorithmischen Entscheidungen betroffen ist, können diese negative Auswirkungen auf sie haben und es gäbe niemanden, der dafür zur Verantwortung gezogen werden könnte, solange die Entscheidung dem Algorithmus zugeschrieben wird.

Es gibt jedoch ebenfalls Autor:innen, die der Meinung sind, dass beim Einsatz von KI-Systemen keine Verantwortungslücke entsteht. So argumentiert zum Beispiel der Technologiephilosoph Johannes Himmelreich, dass ein Mangel an Kontrolle nicht der Grund für die Entstehung von Verantwortungslücken sein kann, solange eine Person die Wahl hat zwischen dem Einsatz oder Nicht-Einsatz eines algorithmischen Systems: „I argue that even if a commander has no control over a certain harm, a commander has control over a risk of harm (…)."[1068] Denn während der Einsatz des algorithmischen Systems mit einem Schadensrisiko verbunden ist, birgt der Nicht-Einsatz selbst nicht dieses Schadensrisiko. Natürlich ist es möglich, dass ohne den Einsatz eines algorithmischen Systems zum Beispiel Anträge nicht schnell genug bearbeitet werden können und somit der aktuelle Zustand auch schon ein Schadensrisiko birgt. Jedoch wird durch den Nicht-Einsatz eines algorithmischen Systems kein weiteres Risiko hinzugefügt. Da für das Tragen der Verantwortung ein gewisses Maß an Kontrolle erforderlich ist, können diejenigen, die über den Einsatz oder Nicht-Einsatz des algorithmischen Systems entscheiden, in angemessener Weise zur Verantwortung gezogen werden.[1069]

Diese widersprüchlichen Einschätzungen hängen damit zusammen, dass der Begriff der Verantwortung, und damit verbunden

1066 Danaher 2022:1
1067 Kiener 2022:584f.
1068 Himmelreich 2019:732
1069 Oimann, Tollon 2024:9; Himmelreich 2019:732f.

die Bedingung der Kontrolle, unterschiedlich definiert werden. Deshalb sind Himmelreich und Köhler der Meinung, dass es Verantwortungslücken nur dann gibt, wenn die Bedingung für Verantwortung eine strenge Form der Kontrolle ist. Denn wenn die verantwortliche Person volitionale Kontrolle über die Handlung und ihre absehbaren Folgen haben muss, ist das in vielen Situationen schwer umsetzbar. Wenn aber die für Verantwortung notwendige Kontrolle als Beaufsichtigung (supervision) verstanden werde, dann gäbe es das Problem der Verantwortungslücken nicht.[1070] Für die Definition von Kontrolle als Beaufsichtigung beziehen sich die Autoren auf Nyholms Argument, dass es sinnvoll sei, die Ergebnisse von Systemen, die von Menschen kontrolliert, aktualisiert oder gestoppt werden können, als Mensch-Maschine-Kollaborationen zu verstehen. Hierbei trifft der Mensch die Entscheidungen, und die maschinellen Kollaborateure stehen unter menschlicher Aufsicht. Die Maschinen agieren unter der Aufsicht der Menschen, so wie Kinder unter der Aufsicht ihrer Eltern agieren. Die beteiligten Menschen sind für das, was die Maschinen tun, deshalb verantwortlich, weil sie diese Mensch-Maschine-Kooperationen initiieren und dann überwachen und steuern – und im Gegensatz zu den Maschinen sind sie zu moralisch verantwortlichem Handeln fähig.[1071]

In der Realität ist es meist jedoch nicht nur ein einzelner Mensch, der die Kooperation initiiert, überwacht und steuert. Deshalb ist trotz dieser die Beziehung klarstellenden Definition noch nicht eindeutig, wer wirklich die Verantwortung trägt. Laut Sparrow besteht die einzig mögliche Lösung darin, die Verantwortung einer geeigneten Person zu übertragen. Eine Person sei dann geeignet für die Zuweisung der Verantwortung für algorithmische Ergebnisse, wenn sie die Verwendung der Algorithmen anordnet.[1072] Kiener hingegen plädiert dafür, dass eine der involvierten Personen aktiv die Verantwortung übernimmt: „It is only the normative power of taking responsibility at will, viz. the power to impose moral answerability on oneself, that could create moral answerability and thereby satisfy the answerability interests of those who have been harmed."[1073] Es

1070 Himmelreich, Köhler 2022
1071 Nyholm 2018:1216f.
1072 Sparrow 2007:74
1073 Kiener 2022:584f.

gibt viele Argumente die dafür sprechen, aktiv Verantwortung zu übernehmen – unter der Bedingung, dass man an der zu verantwortenden Handlung beteiligt war und dementsprechend die Verantwortungsinteressen anderer befriedigen kann.[1074]

Erstens gibt es nicht nur negative Verantwortung, auf Grund derer man für negative Folgen die Schuld trägt oder bestraft wird, sondern es gibt auch positive Verantwortung, auf Grund derer man für positive Folgen gelobt oder belohnt wird.[1075] Zweitens kann es als Tugend verstanden werden, Verantwortung für die eigenen Handlungen und ihre Folgen zu übernehmen, auch wenn man nicht per Definition verantwortlich ist.[1076] Wie bei anderen Tugenden geht es auch bei dieser darum, der richtigen Person zur richtigen Zeit und auf die richtige Weise das richtige Maß an Wiedergutmachung, Entschuldigung oder Schuldgefühl anzubieten.[1077] Drittens kann das Übernehmen von Verantwortung als Ausdruck von Freiheit gelten, weil dies einen freien Willen voraussetzt.[1078] Viertens gibt es pflichtbezogene Gründe, Verantwortung zu übernehmen, weil dies die Möglichkeit bietet, anderen zu helfen, wertvolle moralische Beziehungen zu knüpfen und die eigenen moralischen Fähigkeiten zu entwickeln.[1079] Fünftens kann die Übernahme von Verantwortung laut Kiener sogar selbst verpflichtend sein: „(…) taking responsibility could also be morally *obligatory*, at least in cases in which the harm to another party is severe and there is no one else who could take responsibility in a meaningful way."[1080] Dies könnte zum Beispiel den LKW-Fahrer betreffen, der trotz vorsichtigen Fahrens die Verantwortung für das Überfahren des Kindes übernehmen müsste, weil der angerichtete Schaden so groß ist.[1081]

1074 Kiener 2022:586
1075 Nyholm 2024:195
1076 Betzler, Scherrer 2016:13; Wolf 2001:13
1077 Wolf 2001:13
1078 Sidgwick 2011 [1874]:50
1079 Kiener 2022:588
1080 Ebd.:588
1081 Hier ist nicht außer Acht zu lassen, dass die Eltern ihre Aufsichtspflicht verletzt haben, dass also auch sie tragischerweise in dem Beispiel verantwortlich sind für den Tod ihres Kindes. Trotzdem sollte der LKW-Fahrer die Verantwortung für seinen Anteil an dem Unfall übernehmen.

8.1.3 Verantwortungsvoll handeln

Die bisher thematisierte rückwärtsgerichtete Verantwortungszuordnung behandelt die Frage, wer für etwas Geschehenes verantwortlich gemacht werden kann. Aber es gibt auch vorausschauende Verantwortungsübernahme. Nyholm formuliert das so: „(…) before anything bad happens, there can also be a question of whose responsibility it is to try to make sure that such an outcome does not come about."[1082] Bei vorausschauender (prospektiver) Verantwortung sollte schon im Vorhinein dafür gesorgt werden, dass die verantworteten Handlungen keine negativen Folgen haben, für die man dann die Schuld tragen müsste. Es bestehen also Verpflichtungen zum Risikomanagement, um Schäden vermeiden zu können.[1083] Die vorausschauende Verantwortung beschreibt nicht nur eine Zuständigkeit, sondern involviert eine normative Forderung. So bedeutet Verantwortung tragen nicht nur, dass kein Schaden verursacht werden sollte, sondern auch, dass man sich um das Wohl anderer Menschen kümmern sollte. Wenn also eine Person für das Wohl anderer Menschen verantwortlich war, bedeutet dies, dass die Person bei Schädigung der Menschen zur Rechenschaft gezogen werden kann.[1084] Laut Kiener ist der Unterschied zwischen rückwärtsgerichteter Verantwortung (liability) und vorausschauender Verantwortung (answerability), dass man im Nachhinein als bloßes passives Objekt von Schuld oder Strafe behandelt wird, während man im Vorhinein als rationaler moralischer Akteur angesprochen wird, der eine echte moralische Auseinandersetzung wert ist.[1085] Die vorausschauende Verantwortung muss dem Technologieforscher Richard Owen und seinen Kollegen zufolge Unsicherheiten über die Zukunft berücksichtigen und hat die Ausprägungen Fürsorge und Reaktionsfähigkeit (care and responsiveness). Mit Hilfe der Fürsorge kann schon vor der Entwicklung neuer Technologien darüber nachgedacht werden, was diese (nicht) tun sollen. Reaktionsfähigkeit ermöglicht es,

1082 Nyholm 2024:195
1083 Mieth, Bambauer 2016:7; Himmelreich 2019:733
1084 Coeckelbergh 2016:751; Mieth, Bambauer 2016:7
1085 Kiener 2022:591

während der Entwicklung auf die Ansichten, Perspektiven und Rahmenbedingungen aller Interessengruppen einzugehen.[1086]

Jonas ist ebenfalls der Meinung, dass es sich bei den Handlungsfolgen, für die man verantwortlich ist, auch um Folgen handeln kann, die weit in der Zukunft liegen und womöglich zukünftige Generationen betreffen. Er entwickelt deshalb den folgenden Imperativ: „(...) »Handle so, daß die Wirkungen deiner Handlung verträglich sind mit der Permanenz echten menschlichen Lebens auf Erden« (...)."[1087] Laut Jonas ist es die Pflicht der Menschen, sich eine geistige Art der Furcht vor möglichen negativen Folgen des eigenen Handelns anzueignen. Diese „imaginative(...) Kasuistik"[1088] soll dabei helfen, mögliche zukünftige Risiken heutiger Handlungen zu erkennen. Das heutige Handeln soll sich laut Jonas an solchen möglichen Risiken orientieren, um künftige Katastrophen zu vermeiden.[1089]

In diesem vorausschauenden Sinne bedeutet verantwortlich sein nicht, dass man für ein vergangenes Unrecht schuldig oder haftbar gemacht wird, sondern dass die Akteure ihre Handlungen von vornherein in einer moralisch angemessenen Weise ausführen und dafür sorgen, dass die gewünschten Ergebnisse erzielt werden.[1090] So schreibt Young: „The point is not to compensate for the past, but for all who contribute to processes producing unjust outcomes to work to transform those processes."[1091] Sie bezeichnet diese vorausschauende Verantwortung als eine politische Verantwortung und meint damit, dass sie kollektives Handeln erfordert: „Taking responsibility for structural injustice under this model involves joining with others to organize collective action to reform the structures."[1092] Nicht nur diejenigen, die von ungerechten Strukturen profitieren, sondern insbesondere auch die Personen, die unter diesen Strukturen leiden, können zu einer gemeinschaftlichen Verantwortung

1086 Owen et al. 2013:29, 35f.
1087 Jonas 2020[1979]:36
1088 Ebd.:67
1089 Burckhart, Nielsen-Sikora 2020:180f.; Jonas 2020[1979]:65ff.
1090 Young 2011:104
1091 Ebd.:109
1092 Ebd.:112

aufgerufen werden, sich an Aktionen zur Veränderung dieser Strukturen zu beteiligen.[1093]

Nyholm unterscheidet bei der vorausschauenden Verantwortung zwischen negativer und positiver Verantwortung. Während negative vorausschauende Verantwortung die Vermeidung von zukünftigen Schäden zum Ziel hat, zielt positive vorausschauende Verantwortung darauf ab, gute Ergebnisse herbeizuführen. Laut Nyholm lässt sich das Ziel der positiven vorausschauenden Verantwortung im Kontext von KI-Systemen mit „Value Alignment" gleichsetzen. Dieses sogenannte „Value Alignment" von KI (Werteausrichtung der KI) ist die Idee, dass wir KI-Systeme schaffen sollten, die mit menschlichen Werten, Interessen oder Zielen übereinstimmen.[1094] Diese Werteausrichtung sicherzustellen impliziert eine Verantwortung, dafür zu sorgen, dass gute Ergebnisse erzielt werden.[1095] Die positive vorausschauende Verantwortung wird jedoch sowohl vom gesunden Menschenverstand als auch in einflussreichen Moraltheorien als weniger streng angesehen als negative vorausschauende Verantwortung für die Vermeidung von Schäden. So wird Handeln, das gute Ergebnisse fördert, als etwas betrachtet, das wünschenswert und tugendhaft ist, aber optional. In Kants deontologischer Moraltheorie zum Beispiel gilt das Vermeiden der Verletzung der Würde anderer als strenge Pflicht, während die Förderung des Wohls anderer nur als weite (verdienstliche) Pflicht gilt.[1096] Wenn unklar ist, wer für die Schaffung von KI-Systemen mit guten Ergebnissen vorausschauend verantwortlich ist, es aber den Anschein hat, dass jemand dafür verantwortlich sein sollte, besteht laut Nyholm eine vorausschauende positive Verantwortungslücke.[1097]

Nyholm merkt an, dass es nicht leicht ist, eine solche Lücke zu schließen, da sich die Menschen lieber als Verantwortliche für bereits eingetretene gute Ergebnisse einbringen würden (rückwärtsgerichtet), als die Verantwortung dafür zu übernehmen, dass bestimmte gute Ergebnisse zukünftig erreicht werden (vorausschauend).[1098]

1093 Young 2011:113
1094 Nyholm 2024:193, 199
1095 Ebd.:208
1096 Kant 2008:65ff. [429–430]
1097 Nyholm 2024:200, 208
1098 Ebd.:202

Das liegt daran, dass es kostspielig und mühsam sein kann zu versuchen, gute Ergebnisse herbeizuführen. Nyholm findet zwar Ausnahmen, zum Beispiel bei Eltern, die bereitwillig vorausschauende Verantwortung dafür übernehmen, dass es ihren Kindern gut geht. Aber die Bereitschaft ist nicht immer so groß: „The average corporation does not typically eagerly and willingly take on positive responsibilities to make sure that their products promote good outcomes for other people (...)."[1099] Insbesondere wenn Fremde oder erst zukünftige Generationen von den mühsam erarbeiteten guten Ergebnissen profitieren, sind die Menschen in der Regel viel weniger bereit, dafür Verantwortung zu übernehmen.

Laut Hart hängt die vorausschauende Verantwortung für das Wohl anderer Menschen oft mit einer bestimmten sozialen Rolle zusammen: „(...) whenever a person occupies a distinctive place or office in a social organisation, to which specific duties are attached to provide for the welfare of others or to advance in some specific way the aims or purposes of the organization, he is properly said to be responsible for the performance of these duties, or for doing what is necessary to fulfil them."[1100] Eine solche Rolle kann einem Menschen auf Grund bestimmter Eigenschaften zufallen, aber auch, weil ihm die Aufgabe durch eine Vereinbarung oder auf andere Weise zugewiesen wurde.[1101] So sind zum Beispiel Eltern, allein auf Grund ihrer sozialen Rolle, verantwortlich für ihre Kinder – rückwirkend für die Handlungen ihrer Kinder und vorausschauend für deren Wohl.[1102] Auch Führungskräfte in einem Unternehmen sind auf Grund ihrer sozialen Rolle für das Wohl ihrer Mitarbeitenden zuständig, können aber auch für absehbare Folgen der Handlungen von Mitarbeitenden zur Verantwortung gezogen werden. Die Verantwortung wird hingegen zugewiesen, wenn zum Beispiel zwei Personen vereinbaren, dass auf einer Wanderung die eine Person für die Versorgung und die andere Person für die Navigation zuständig ist.[1103]

Verantwortungsvoll handeln heißt, schon im Vorhinein zu bedenken, welche positiven und negativen Folgen die Handlung haben kann und durch frühzeitige Entscheidungen die negativen Folgen

1099 Nyholm 2024:203
1100 Hart 1968:212
1101 Ebd.:213
1102 Mieth, Bambauer 2016:5; Hart 1968:212
1103 Hart 1968:212

zu vermeiden und die positiven Folgen aktiv herbeizuführen. Im Kontext eines Datenökosystems vorausschauend verantwortungsvoll zu handeln bedeutet dementsprechend, dass schon vor Entwicklung und Einführung der Schutz sowohl der Privatsphäre von Datenquellen als auch die Transparenz für die Datennutzenden bedacht werden.

Laut Hart handelt eine Person dann verantwortungsvoll, wenn sie ihre Pflichten ernst nimmt.[1104] Die Begriffe Verantwortung und Pflicht sind nicht gleichbedeutend, und es ist auch nicht einer auf den anderen reduzierbar. Vielmehr handelt es sich laut Mieth und Bambauer um zwei unterschiedliche Konzepte mit gewissen Schnittmengen: „Moralische Verantwortung zielt im Unterschied zu Pflichten auf bestimmte Ereignisse oder Zustände ab, während Pflichten bestimmte Handlungen gebieten (...)."[1105] So ist eine Führungskraft für das Wohl der Mitarbeitenden verantwortlich und hat gleichzeitig gewisse Pflichten den Mitarbeitenden gegenüber, wie zum Beispiel das Einhalten von Sicherheitsregeln am Arbeitsplatz. Laut Mieth und Bambauer können „(...) moralische Pflichten als konkrete praktische Forderungen fungieren (...), die plausibel auf eine zugrundeliegende Verantwortung rückführbar sind."[1106] Das heißt also, dass eine Pflicht aus einer Verantwortung hergeleitet werden kann. Somit geht die Verantwortung über einzelne Pflichten hinaus, aber das Erfüllen von Pflichten kann tatsächlich konstitutiv sein, um der Verantwortung gerecht zu werden.

Aus der Verantwortung für einen guten Umgang miteinander lässt sich somit eine Pflicht, das Wohl anderer zu fördern, herleiten – wenn dies auch im Sinne Kants nur eine weite (verdienstliche) Pflicht ist. Diese Pflicht muss zumindest in bestimmten sozialen Rollen erfüllt werden, um der positiven vorausschauenden Verantwortung gerecht zu werden. Die Bereitschaft einzelner Personen, freiwillig diese Verantwortung zu übernehmen, mag, wie Nyholm befürchtet, gering sein. Aber im Folgenden wird klar, dass es in der Realität sowieso oft nicht nur eine einzelne Person ist, welche die positive oder negative vorausschauende Verantwortung trägt.

1104 Hart 1968:213
1105 Mieth, Bambauer 2016:11
1106 Ebd.:11

8.1.4 Kollektive Verantwortung

Dazu soll an dieser Stelle betrachtet werden, ob es möglich ist, dass mehrere Menschen gemeinsam Verantwortung tragen. Die feministische Ethikerin Tracy Isaacs definiert kollektive Verantwortung folgendermaßen: „An dieser Stelle wird kollektive Verantwortung als die Löblichkeit oder Schuldigkeit, die kollektiven Akteuren für ihre Handlungen gebührt, betrachtet."[1107] Dabei muss eine Gruppe intentional – also mit Gründen – handeln können, um als Akteur zu gelten. Im Fall eines Unternehmens wird dem Philosophen Peter French zufolge dann auf kollektiver Ebene intentional gehandelt, wenn auf Grundlage von unternehmensinternen Entscheidungsstrukturen gehandelt wird.[1108] Diese Strukturen bestehen French zufolge aus einem Organisations- oder Verantwortungsdiagramm, das die Machtstrukturen des Unternehmens abbildet sowie Anerkennungsregeln für Entscheidungen in Form einer Unternehmenspolitik. Damit seien Unternehmen hochgradig strukturierte kollektive Akteure, die als Organisationen bezeichnet werden können – genau wie zum Beispiel Regierungen, Institutionen oder Sportmannschaften.[1109] Es gibt aber auch andere Gruppen, die intentional handeln und als Akteure gelten können – Isaacs nennt sie zielorientierte Kollektive. In einem solchen Kollektiv arbeiten Menschen zusammen, um ein gemeinsames Ziel zu erreichen, das sie nicht ohne Koordination erreichen können. So könnten zum Beispiel Nachbarn zusammenarbeiten, um die gemeinsamen Wege zu reinigen.[1110]

Es gibt verschiedene Positionen dazu, wer bei kollektivem Handeln die Verantwortung für die Handlungen und ihre absehbaren Folgen trägt. Isaacs und French zum Beispiel sind der Meinung, „(...) dass kollektive Akteure Handlungen vollziehen, für die ausschließlich sie selbst rechenschaftspflichtig sind."[1111] Es wirken zwar Individuen an den kollektiven Handlungen mit, aber die einzelnen Mitglieder handeln nicht wie das Kollektiv, dessen Teil sie sind. Manche Handlungen können aus korporativen Gründen entstehen.

1107 Isaacs 2016:5
1108 French 1979:212, 215
1109 Ebd.; Isaacs 2016:5f.
1110 Isaacs 2016:6
1111 Ebd.:8

Diese korporativen Gründe unterscheiden sich qualitativ von den persönlichen Gründen, welche die Mitglieder für ihr individuelles Handeln haben.[1112] Unternehmen haben solch korporative Gründe, weil sie ein Interesse daran haben die Dinge zu tun, die wahrscheinlich zur Verwirklichung ihrer festgelegten Unternehmensziele führen, unabhängig von den flüchtigen Eigeninteressen der Direktor:innen oder Manager:innen. Dabei liegt es trotzdem im Interesse eines Unternehmens, dass seine Mitglieder die Unternehmensziele als hilfreich für die Erreichung ihrer eigenen Ziele ansehen.[1113] Da es aber der kollektive Akteur ist, der handelt, muss laut Isaac auch eben diesem Kollektiv die Verantwortung für die Handlung und ihre Folgen zugeschrieben werden: „Insofern kollektive Akteure auf Basis ihrer eigenen Gründe intentional handeln, können wir sie als kollektive Akteure betrachten, die für ihre Handlungen Lob oder Tadel verdienen."[1114]

Viele Autor:innen zweifeln daran, dass kollektive Akteure Verantwortung tragen können, da sie im Fall von negativen Handlungsfolgen nicht angemessen bestraft werden können und deshalb die beteiligten Individuen entweder zu Unrecht zu viel Verantwortung aufgelastet bekommen oder sich ihrer Verantwortung entziehen. Stattdessen müssten dann Entitäten, die nicht zum intentionalen Handeln fähig sind, die Verantwortung tragen.[1115] Darum gibt es den alternativen Ansatz der individualistischen Konzeption kollektiver Verantwortung, der besagt, dass wir „(...) jede Zuschreibung kollektiver Verantwortung auf ein Set von Aussagen über die Verantwortung einzelner menschlicher Akteure reduzieren können."[1116] Die Idee ist, dass sich kollektive Handlungen auf individuelle Intentionen reduzieren lassen, sodass jedes (handelnde) Individuum einen Teil der Verantwortung für die Folgen der kollektiven Handlung trägt.[1117] Es gibt aber auch den radikaleren Ansatz, der kollektive Verantwortung vollkommen verwirft. Die Begründung ist, dass niemand bestraft

1112 French 1979:215
1113 Ebd.:214
1114 Isaacs 2016:9
1115 Ebd.:14–18
1116 Ebd.:10
1117 Kutz 2000:85; Isaacs 2016:10f.

werden kann, wenn sich kollektive Handlungen nicht auf individuelle Handlungen reduzieren lassen.[1118]

Isaacs ist der Meinung, dass bei diesen Einwänden von falschen Voraussetzungen ausgegangen wird. So wird die kollektive Verantwortung weder auf die einzelnen Mitglieder des Kollektivs verteilt, noch wird durch die kollektive Verantwortung ausgeschlossen, dass die einzelnen Mitglieder für ihr Mitwirken an den kollektiv erwirkten Folgen Verantwortung tragen können: „Kollektive Akteure bleiben für ihre kollektiven Handlungen verantwortlich, aber die individuellen menschlichen Akteure können immer noch für ihre Mitwirkungen an diesen Handlungen verantwortlich sein."[1119] Isaacs verteidigt zudem die Auffassung, dass auch Kollektive zu intentionalem Handeln fähig sind und dementsprechend selbst als Akteure gelten können. Denn ihrer Meinung nach müssen Intentionen nicht als mentale Zustände aufgefasst werden, welche ein Kollektiv nicht aufweisen könnte. Vielmehr plädiert Isaacs für eine funktionale Konzeption von Intention, sodass auf Seiten der kollektiven Akteure kein Bewusstsein vorausgesetzt wird.

Aus ihrer Sicht ist es zudem möglich, dass kollektive Akteure bestraft werden, indem zum Beispiel Staaten geteilt und Unternehmen zerschlagen oder mit einer Geldstrafe belegt werden. Eine solche Strafe für das Kollektiv hat dabei auch einen Effekt auf seine Mitglieder, aber das hält Isaacs für unproblematisch: „Die unschuldigen Individuen werden nicht im engeren Sinne bestraft für das Fehlverhalten der Kollektive, denen sie angehören."[1120] Es sei eine unglückliche Folge der Strafe, wenn unschuldige Menschen ihre Konsequenzen zu spüren bekämen, aber dies sei kein Grund dafür, den Schuldigen nicht zu bestrafen. Die kollektive Verantwortung sei unabdingbar, sie erfülle eine sehr wichtige Rolle bei der Verfolgung von sozialen Ungerechtigkeiten – wie zum Beispiel Genoziden – da keine Einzelperson dafür verantwortlich gemacht werden könne. Außerdem sei es, insbesondere in Bezug auf die vorausschauende Verantwortung, wichtig zu begreifen, dass koordiniertes Vorgehen als Kollektiv notwendig ist, um Veränderungen zu erreichen.[1121]

1118 Narveson 2002:185; Isaacs 2016:12f.
1119 Isaacs 2016:16
1120 Ebd.:15
1121 Ebd.:16–21

Hier würde Young zustimmen. Sie fasst den Begriff der Verantwortung in ein Modell der sozialen Bindung (social connection model): „The social connection model finds that all those who contribute by their actions to structural processes with some unjust outcomes share responsibility for the injustice."[1122] Im Rahmen des Modells der sozialen Beziehungen wird Verantwortung im Wesentlichen geteilt und kann daher nur durch kollektives Handeln wahrgenommen werden.[1123] Laut Young bedeutet geteilte Verantwortung, dass man persönlich Verantwortung trägt, aber man trägt sie nicht alleine, sondern in dem Bewusstsein, dass andere sie mit einem tragen. Dadurch, dass man diese geteilte Verantwortung anerkennt, erkennt man auch an, dass man Teil eines Kollektivs ist, das gemeinsam Ungerechtigkeiten produziert. Die Verantwortung wird deshalb mit anderen geteilt, weil die Schäden von vielen gemeinsam, im Rahmen akzeptierter Institutionen und Praktiken, verursacht werden. Dabei ist kein Mitglied des Kollektivs in der Lage genau zu bestimmen, welcher Teil seiner/ihrer Handlungen zu welchem Aspekt der Ungerechtigkeit führt.[1124]

Geteilte Verantwortung ist dabei nicht das Gleiche wie kollektive Verantwortung. Denn während ein Kollektiv von Personen, wie zum Beispiel ein Unternehmen, für negative Folgen verantwortlich gemacht werden kann, ohne dass eine der Einzelpersonen, aus denen es besteht, dafür verantwortlich ist[1125], bedeutet geteilte Verantwortung, dass alle Einzelpersonen in einer Gruppe eine persönliche Verantwortung für negative Folgen tragen, die von der Gruppe hervorgebracht werden.[1126]

Im Folgenden stellt sich nun die Frage, wie sich die theoretischen Überlegungen zur Verantwortung in die Praxis umsetzen lassen. Es soll geklärt werden, wer im Datenökosystem Verantwortung trägt und ob es sich um ein Individuum oder ein Kollektiv handelt, welches sowohl mögliche negative Folgen verantworten muss als auch vorausschauende Verantwortung für die zukünftige Gestaltung trägt.

1122 Young 2011:96
1123 Ebd.:105
1124 Ebd.:109f.
1125 Juristisch trägt dann je nach Festlegung für das Kollektiv ‚Unternehmen' wiederum ein Geschäftsführer oder Vorstandsvorsitzender die Verantwortung.
1126 Young 2011:110

8.2 Verantwortliche im Datenökosystem

In einem Datenökosystem muss einerseits rückwirkend für Schäden Verantwortung übernommen werden und andererseits vorausschauend für eine möglichst gute Gestaltung des Systems. Es muss jemand dafür sorgen, dass niemandem unzulässigerweise ein Risiko auferlegt wird, dass Datenquellen nicht die Kontrolle über ihre Informationen verlieren, und dass Datennutzende nicht durch Verzerrungen und Intransparenz zu falschen Überzeugungen gelangen. Im besten Fall wird sogar vorausschauend die Förderung von Privatsphäre und Transparenz verantwortet. Sollten Datenquellen die Kontrolle über ihre Informationen verlieren und Datennutzende durch Verzerrungen und Intransparenz zu falschen Überzeugungen gelangen, muss jemand dafür geradestehen. Bei der ethischen Gestaltung eines Datenökosystem stellt sich also die Frage, wer verantwortlich ist.

Im AI Act ist eine der Bedingungen für den Einsatz von Hochrisiko-KI-Systemen die Beaufsichtigung durch natürliche Personen[1127] – nur so kann die Verantwortung bei einem menschlichen Akteur liegen. Sowohl das Digitale-Dienste-Gesetz (DDG) als auch der AI Act halten die Anbieter von Vermittlungsdiensten bzw. von KI-Systemen für verantwortlich: „Enthält ein Produkt ein KI-System, (...) so sind die Anbieter dafür verantwortlich, sicherzustellen, dass ihr Produkt alle geltenden Anforderungen (...) vollständig erfüllt."[1128] Artikel 16 des AI Acts schreibt den Anbietern von Hochrisiko-KI-Systemen Pflichten zu. Ebenso thematisiert Artikel 23 die Pflichten der Einführer, die ein Hochrisiko-KI-System in Verkehr bringen und Artikel 24 die Pflichten der Händler, die ein Hochrisiko-KI-System auf dem Markt bereitstellen. Laut Artikel 25 können Händler, Einführer, Betreiber oder sonstige Dritte als Anbieter eines Hochrisiko-KI-Systems gelten. Betreiber sind dabei diejenigen, die ein KI-System in eigener Verantwortung und im Rahmen einer beruflichen Tätigkeit verwenden.[1129] Es können jedoch auch notifizierte Stellen die Verantwortung für Hochrisiko-KI-Systeme übernehmen.[1130]

1127 Art. 14, Abs. 1, Satz 1 EU AI Act
1128 Art. 8, Abs. 2, Satz 1 EU AI Act
1129 Art. 3, Abs. 4 EU AI Act
1130 Art. 36, Abs. 3, Satz 2 EU AI Act

Wenn sich Risikoexponierte nicht freiwillig dem Risiko aussetzen, sondern sie durch bestimmte Umstände dazu gezwungen sind, liegt laut Hermansson und Hansson die Verantwortung für die Vermeidung des Schadeneintritts und die Erreichung von guten Ergebnissen bei denjenigen, die das Risiko und den potenziellen Schaden verursachen.[1131] Wenn jedoch Kinder oder Maschinen eine Rolle bei der Erzeugung eines Ergebnisses spielen, kann es sich laut Sparrow nur um eine kausale und nicht um eine moralische Rolle handeln, da Kinder und Maschinen nicht autonom genug sind: „(...) if anyone is responsible for that outcome, it is the person who placed them in the position where they played that causal role."[1132] Es sind die Personen für das Ergebnis verantwortlich, welche die Kinder oder die Maschinen in die Situation versetzt haben, eine kausale Rolle bei der Erzeugung des Ergebnisses zu spielen. Auch Nyholm vergleicht die Handlungsfähigkeit eines kleinen Kindes mit der einer autonomisierten Maschine: Sowohl zu Kindern als auch zu Maschinen hat die verantwortliche Person eine Beziehung der Beaufsichtigung (supervision). Er ist der Meinung, dass autonomisierte Maschinen als Teil von Mensch-Maschine-Teams betrachtet werden sollten, in denen die beteiligten Menschen die Rolle von Manager:innen oder Beaufsichtigenden spielen. Die Menschen tragen also die Verantwortung für die Maschinen, wie eine Führungsposition innerhalb eines hierarchischen Teams die Verantwortung für das Team trägt.[1133]

Es stellt sich trotzdem die konkrete Frage, welche Menschen oder welche Gruppe von Menschen die Verantwortung tragen. Hierzu gibt es insbesondere in Bezug auf die Verantwortung für autonomisierte Maschinen eine lebendige Diskussion in der Literatur. Die Philosophen mit dem Schwerpunkt politischer Theorie Alexander Hevelke und Julian Nida-Rümelin zum Beispiel argumentieren, dass die Hersteller von KI-Systemen nicht für deren negative Folgen verantwortlich gemacht werden sollten, da dies die Hersteller von der Entwicklung selbstlernender Maschinen abhalten könnte. Es ließe sich aber moralisch dafür argumentieren, dass die Entwicklung von KI-Systemen gefördert werden sollte, zumindest wenn ihre Ein-

1131 Hermansson, Hansson 2007:142
1132 Sparrow 2007:74
1133 Nyholm 2018:1210; Nyholm 2024:204

führung Schaden verringern kann und kein erhöhtes Risiko bedeutet.[1134]

Es gibt also pragmatische Gründe, die gegen eine Verantwortungsübernahme durch die Hersteller sprechen. Dies bedeutet aber laut Nyholm nicht zwangsläufig, dass eine solche Verantwortungsübernahme an sich nicht gerechtfertigt ist.[1135] Ich würde sogar argumentieren, dass es im Gegenteil sinnvoll wäre, die Hersteller der KI-Systeme zur Verantwortung zu ziehen und dadurch zum tieferen Nachdenken anzuregen. Sie sollen dadurch nicht von der Entwicklung der Maschinen abgehalten werden, aber sie sollen sich schon fragen, ob sie die Erfindung und den Einsatz solcher Maschinen verantworten können. Die Übernahme der Verantwortung wäre auch ein Hinweis darauf, dass Hersteller davon überzeugt sind, dass ihr Produkt keinen erheblichen Schaden anrichten wird, sondern eher einen Nutzen mit sich bringt.

Hevelke und Nida-Rümelin sind zudem der Meinung, dass die Menschen, welche autonomisierte Maschinen betreiben, nicht für deren negative Folgen verantwortlich gemacht werden sollten, da es nur von Glück oder Pech abhänge, ob die Maschine, die man betreibt, negative Folgen hervorbringt oder nicht. Alle Personen, die sich in ein selbstfahrendes Auto setzen und damit zu Betreibern einer automatisierten Maschine werden, verursachen nach Meinung der Autoren in gleicher Weise Risiken – es sei nur auf Pech zurückzuführen, wenn das betriebene Auto einen Unfall verursacht.[1136] Nyholm hingegen argumentiert, dass es gerade die Betreibenden von autonomisierten Maschinen sind, welche die Verantwortung für deren negative Folgen tragen sollten: "(…) given that the human operator here collaborates with the car in the role as a sort of an active supervisor, it makes sense to view the human party to this collaboration as being the responsible party."[1137] Jemand, der ein selbstfahrendes Auto betreibt, hat die Kontrolle, dieses jederzeit zu stoppen und hat deshalb auch die Verantwortung das Auto anzuhalten, bevor negative Folgen auftreten.

1134 Hevelke, Nida-Rümelin 2015:620–623
1135 Nyholm 2018:1205, Fußnote 4
1136 Hevelke, Nida-Rümelin 2015:627f.
1137 Nyholm 2018:1212

Wenn allerdings das selbstfahrende Auto unter der genauen Beobachtung der Designer und Hersteller steht, die eingreifen und die Hard- und Software ihres Produktes aktualisieren, wenn sie es für notwendig erachten, sind sie eine Schlüsselpartei in der Mensch-Roboter-Zusammenarbeit und tragen laut Nyholm die Hauptverantwortung dafür, wie das Auto funktioniert. „(...) the traveller sets the end, but (...) the car company determines the means by which that end is achieved."[1138] Das selbstfahrende Auto kollaboriert also auf verschiedene Weise mit unterschiedlichen Menschen. Denn das Auto hilft den Betreibenden dabei, ein Reiseziel zu erreichen – gleichzeitig haben die Betreibenden die Kontrolle, das Auto zu stoppen. Zudem besteht zwischen dem Auto und den Herstellenden eine kollaborative Beziehung, weil das Auto für die Herstellenden eine Dienstleistung erbringt – dabei kontrollieren die Herstellenden, auf welche Weise die Dienstleistung erbracht wird.[1139] Hier gibt es also zwei Parteien, welche gleichzeitig die Rolle der Beaufsichtigenden einnehmen. Wenn das selbstfahrende Auto ein Fahrrad nicht erkennen kann und deshalb anfährt, sind die Herstellenden für den systemischen Fehler verantwortlich, aber die Betreibenden des Autos sind verantwortlich dafür, nicht rechtzeitig gebremst zu haben. Die Verantwortung reicht in diesem Fall so weit wie der jeweilige Kontrollbereich.

Verantwortung kann aber auch die Notwendigkeit umfassen, sich anderweitig Informationen einzuholen. Insbesondere wenn man mit im Datenökosystem verarbeiteten Daten umgeht und diese für Datennutzende aufbereitet oder auf deren Grundlage weitreichende Entscheidungen trifft, ist es ein Teil der Verantwortung, sich anderweitig zu informieren, um die Richtigkeit der Ergebnisse überprüfen zu können. Laut Buijsman und Veluwenkamp ist man sogar schuld an einem Fehler, wenn man den algorithmischen Ergebnissen blind gefolgt ist, obwohl widersprechende Informationen leicht verfügbar waren. Leichte Verfügbarkeit bedeutet insbesondere, dass genügend Zeit für die Aufbereitung der Daten oder das Fällen der Entscheidung zur Verfügung steht.[1140]

1138 Nyholm 2018:1214
1139 Ebd.:1214
1140 Buijsman, Veluwenkamp 2023:543

Für vorausschauende Verantwortung ist es kein Problem, dass mehrere Parteien dafür verantwortlich sind, zukünftig Schaden zu vermeiden und gute Ergebnisse zu erzielen. Aber in Bezug auf die rückwärtsgerichtete Verantwortung stellt sich die Frage, welche Partei im Schadensfall die Verantwortung übernimmt. Hevelke und Nida-Rümelin schlagen als Lösung vor, dass alle, die selbstfahrende Autos nutzen, eine Steuer oder eine Pflichtversicherung zahlen, die teilweise auf der Anzahl der jährlich gefahrenen Kilometer basiert – davon soll dann durch selbstfahrende Autos entstandener Schaden beglichen werden.[1141] Damit beantworten sie aber nur die Frage, wer für einen Schaden haftet, und nicht, wer dafür verantwortlich ist, diesen zu vermeiden. Danaher fügt hinzu, dass die Menschen im Schadensfall jemanden suchen, der die Verantwortung übernimmt und der eine Strafe verdient hat. Wenn es aber niemanden gibt, dem die Verantwortung zugeschrieben werden kann, entsteht eine Vergeltungslücke (retribution gap).[1142] Laut Nyholm kann diese Vergeltungslücke nicht durch eine Steuer oder eine Pflichtversicherung geschlossen werden, da Menschen einen starken Impuls haben, im Schadensfall jemanden zu bestrafen.[1143]

Im Datenökosystem tragen diejenigen Verantwortung für die Risikoauferlegung, die entweder Kontrolle über die betreffenden Handlungen und ihre absehbaren Folgen haben[1144], eine beaufsichtigende Rolle im Sinne Nyholms[1145] einnehmen, oder durch ihre soziale bzw. zugeschriebene Rolle dazu verpflichtet sind, die Verantwortung zu übernehmen[1146]. Datengenerierende haben im Datenökosystem die Kontrolle darüber, ob und wie sie den Datenquellen das Risiko der Verletzung ihrer Privatsphäre auferlegen. Sollten unvorhersehbare Folgen auftreten und die Daten der Datenquellen zum Beispiel von Unbefugten zweckentfremdet werden, hatten die Datengenerierenden trotzdem auf Grund ihrer Rolle im Datenökosystem eine beaufsichtigende Rolle und sind weiterhin verantwortlich, weil sie

1141 Hevelke, Nida-Rümelin 2015:627
1142 Danaher 2016:299, 302
1143 Nyholm 2018:1205f.
1144 Frankfurt 1971; Fischer, Ravizza 1998; Kane 1996; Smith 2005
1145 Nyholm 2018:1216
1146 Hart 1968:216

die Daten gesammelt und gespeichert haben. Gleichermaßen haben Datenverarbeitende im Datenökosystem die Kontrolle darüber, ob und wie sie die Daten verarbeiten und die Ergebnisse aufbereiten und dementsprechend den Datennutzenden das Risiko der fehlenden Transparenz auferlegen. Sollten unvorhersehbare Folgen auftreten, hatten die Datenverarbeitenden trotzdem auf Grund ihrer Rolle im Datenökosystem eine beaufsichtigende Rolle und sind weiterhin verantwortlich, weil sie die Daten verarbeitet und die Ergebnisse bereitgestellt haben. Deshalb ist das Erfüllen der Pflichten zum Schutz der Privatsphäre und zur Ermöglichung der Transparenz konstitutiv, um der Verantwortung gerecht zu werden.[1147]

Im Folgenden wird ausführlicher betrachtet, wer bei der Gestaltung eines Datenökosystems die Verantwortung für den Schutz der Privatsphäre von Datenquellen und die Bereitstellung der Transparenz für Datennutzende trägt. Da den Datenquellen und den Datennutzenden zwei verschiedene Risiken von unterschiedlichen Akteuren auferlegt werden, werden diese separat betrachtet.

8.2.1 Verantwortung für die Privatsphäre der Datenquellen

Verantwortlich für den vorausschauenden Schutz der Privatsphäre der Datenquellen einerseits und die Vermeidung von Verletzung der Privatsphäre der Datenquellen andererseits sind diejenigen, die in einer Relation zu den entsprechenden Handlungen und ihren Folgen stehen. Somit werden Akteure benötigt, die entweder Kontrolle über die betreffenden Handlungen und ihre absehbaren Folgen haben[1148], eine beaufsichtigende Rolle im Sinne Nyholms[1149] einnehmen, oder durch ihre soziale bzw. zugeschriebene Rolle dazu verpflichtet sind, die Verantwortung zu übernehmen[1150].

Es stellt sich nun die Frage, welche Handlungen potenziell positive oder negative Folgen für die Privatsphäre der Datenquellen haben könnten und dementsprechend verantwortet werden müssen. Im Datenökosystem ist dies die Handlung der Datensammlung und

1147 Hart 1968:11
1148 Frankfurt 1971; Fischer, Ravizza 1998; Kane 1996; Smith 2005
1149 Nyholm 2018:1216
1150 Hart 1968:216

-speicherung und insbesondere der fehlenden diesbezüglichen Informationsweitergabe an die Datenquellen. Gesucht sind also diejenigen, die im Kontext eines Datenökosystems Daten sammeln und speichern und für das Informieren der Datenquellen zuständig sind bzw. diejenigen, die diesen Prozess beaufsichtigen. In meinem Konzept zum Datenökosystem bezeichne ich die gesuchte Personengruppe als „Datengenerierende".

Datengenerierende sind auch diejenigen, die im Fall negativer Folgen die Schuld zugeschrieben bekommen und im Fall positiver Folgen ein Lob verdienen. Diese Schuldzuschreibung ist allerdings abhängig davon, ob es sich um einen klaren kausalen Zusammenhang zwischen Handlung und Folgen handelt, oder ob viele ähnliche, durch soziale Strukturen bedingte, Handlungen zu der negativen Folge sozialer Ungerechtigkeit beitragen. Denn laut Young kann man im letzteren Fall den einzelnen Handelnden keine Schuld zuschreiben.[1151] An dieser Stelle fragt sich, ob in einem Datenökosystem geteilte oder kollektive Verantwortung für die Privatsphäre der Datenquellen getragen wird.

Betrachten wir zuerst die von Young präferierte geteilte Verantwortung, bei welcher es nicht möglich ist genau zu bestimmen, was in den einzelnen Handlungen zu welchen Aspekten einer Ungerechtigkeit führt.[1152] In Bezug auf die Privatsphäre im Datenökosystem ist es relevant, dass die gleiche Handlung sehr häufig stattfindet, wenn viele verschiedene Akteure Daten sammeln und speichern. Dies erhöht die Wahrscheinlichkeit, dass Individuen, deren Daten gesammelt und gespeichert werden, die Kontrolle über Zugang zu und Nutzung ihrer Daten verlieren und dass somit ihre Privatsphäre verletzt wird. Lange haben akzeptierte Institutionen und Praktiken gefehlt, und sie fehlen in vielen Ländern auch heute noch. Aber in der Europäischen Union gibt es mit der Datenschutzgrundverordnung eine Einigung darüber, welcher Umgang mit personenbezogenen Daten akzeptiert wird. Es muss beobachtet werden, ob die Handlungen auf dieser Grundlage zu sozialen Ungerechtigkeiten führen. Das könnte dadurch geschehen, dass nur bestimmte soziale Gruppen die Kontrolle über ihre Daten verlieren, weil diese von besonderem Interesse sind. So können sich diejenigen, welche die Da-

1151 Young 2011:111
1152 Ebd.:109f.

ten sammeln und speichern, an alle akzeptierten Praktiken halten, aber es können andere Interessenten diese Sammlung missbrauchen, um Informationen über bestimmte soziale Gruppen einzuholen. Solange aber eine solche soziale Ungerechtigkeit nicht als Folge vieler Handlungen im Datenökosystem festgestellt wird, ist es nicht sinnvoll, hier von geteilter Verantwortung zu sprechen.

Kollektive Verantwortung hingegen ist ein hilfreiches Konzept im Kontext von Privatsphäre in Datenökosystemen, da hier selten Handlungen einzelner Personen zum Ergebnis führen, sondern kollektive Akteure, wie zum Beispiel Unternehmen, eine Rolle spielen. Ein Unternehmen kann dann als kollektiver Akteur gelten, wenn intentional, auf Grundlage von unternehmensinternen Entscheidungsstrukturen, gehandelt wird. Hierfür ist es laut French notwendig, dass es sowohl ein Organisations- oder Verantwortungsdiagramm gibt, welches die Machtstrukturen des Unternehmens abbildet, als auch Anerkennungsregeln für Entscheidungen gelten. Es wirken zwar Individuen an den kollektiven Handlungen von Unternehmen mit, „(...) aber nur der kollektive Akteur führt die gesamte Handlung auf der kollektiven Ebene aus."[1153] Wenn also das Unternehmen als kollektiver Akteur handelt, kann diesem Kollektiv die Verantwortung für die Handlung und ihre Folgen zugeschrieben werden.[1154]

Datengenerierende sind für das Sammeln und Speichern der Daten und für das Informieren der Datenquellen zuständig, und sie können dabei als kollektiver Akteur verstanden werden. Diese Rolle umfasst in der Praxis viele verschiedene Aufgaben auf allen organisatorischen Ebenen eines Unternehmens. Die Entscheidung, welche personenbezogenen Daten gesammelt und gespeichert werden und inwiefern die Datenquellen darüber informiert werden, treffen Menschen auf der Managementebene, aber die Anforderungen an das Produkt werden auf der Ebene des Concept Development definiert, und die technische Umsetzung dieser Vorgaben erfolgt dann im Development Process.[1155] Solange aber Entscheidungen zur Datensammlung und -speicherung und zum Informieren der Datenquellen von einem kollektiven Akteur auf Grundlage von unternehmensinternen Entscheidungsstrukturen getroffen werden, trägt eben

1153 Isaacs 2016:9
1154 French 1979:212ff.; Isaacs 2016:5–9
1155 Gogoll et al. 2021:1089

dieser kollektive Akteur die Verantwortung. Es ist nicht möglich, generell festzustellen, welche einzelnen Mitglieder zur Gruppe der Datengenerierenden gehören und dementsprechend einen Einfluss haben auf die Privatsphäre der Datenquellen. Es ist die Aufgabe jedes Unternehmens, sich entsprechend der bei ihnen konkret beteiligten Gruppen über die eigenen Verantwortlichkeiten bewusst zu werden.

8.2.2 Verantwortung für die Transparenz für Datennutzende

Verantwortlich für die vorausschauende Förderung der Transparenz für die Datennutzenden einerseits und die Vermeidung von falschen Überzeugungen der Datennutzenden auf Grundlage der zur Verfügung gestellten Informationen andererseits sind diejenigen, die in einer Relation zu den betreffenden Handlungen und ihren Folgen stehen. Bei den Handlungen, die potenziell positive oder negative Folgen für die Transparenz im Datenökosystem haben könnten und die dementsprechend verantwortet werden müssen, handelt es sich um das Verarbeiten der Daten mit Hilfe von Algorithmen sowie um die Aufbereitung der Ergebnisse für die Datennutzenden. Diejenigen, die im Datenökosystem die Datenverarbeitung und -aufbereitung beaufsichtigen, sind für die absehbaren Folgen verantwortlich.[1156] In meinem Konzept zum Datenökosystem nenne ich die gesuchte Personengruppe „Datenverarbeitende".

Hier stellt sich die Frage, ob durch die Datenverarbeitenden geteilte oder aber kollektive Verantwortung für die Transparenz für die Datenquellen getragen wird. Auch hier ist es nur sinnvoll von geteilter Verantwortung zu sprechen, wenn vielfache Datenverarbeitung im Rahmen akzeptierter Institutionen und Praktiken soziale Ungerechtigkeit zur Folge hat. Durch verschiedene Verzerrungen in den Daten, die verarbeitet werden, ist dies tatsächlich nicht unwahrscheinlich[1157] – das wird zum Beispiel an dem Fall von Amazon deutlich.[1158]

1156 Nyholm 2018:1216
1157 Friedman, Kahn 2008:1253
1158 Friedman, Nissenbaum 1996:332

Laut Friedman und Nissenbaum liegt die Verantwortung für die Vermeidung oder Minimierung von Verzerrungen bei denjenigen, die das System gestalten, weil nur sie dabei darauf achten können, dass Vorurteile nicht reproduziert werden, nur sie bei der Gestaltung schon die Nutzung des Systems antizipieren und einen Nutzungskontext festlegen können.[1159] Zudem schlage ich in dieser Arbeit vor, dass die Ergebnisse im zum Verständnis notwendigen Kontext dargestellt werden sollten, sodass die Datennutzenden nachvollziehen können, in welchem Kontext das Ergebnis zu verstehen ist. Auf diese Weise kann auch das Ergebnis zum Beispiel eines Berufsempfehlungs-Algorithmus eingeordnet werden, wenn deutlich gemacht wird, dass der Algorithmus mit seiner Datengrundlage aus der Vergangenheit nur reproduzieren kann, was in der Vergangenheit mit den gegenwärtigen Qualifikationen und Voraussetzungen möglich war. Eine Voraussetzung für diese Art von Transparenz ist natürlich, dass diejenigen, die die Ergebnisse im zum Verständnis notwendigen Kontext darstellen sollen, selbst verstehen, wie das Ergebnis zustande gekommen ist (nach welchen Kriterien der Input zum Output wurde).

Es ist also in bestimmten Fällen möglich, dass sich die für Transparenz Verantwortlichen die Verantwortung für das Erzeugen sozialer Ungerechtigkeit teilen. Trotzdem ist die kollektive Verantwortung auch in Bezug auf Transparenz für die Datennutzenden ein hilfreiches Konzept, da auch hier selten die Handlungen einzelner Personen direkt zum Ergebnis führen, sondern kollektive Akteure, wie zum Beispiel Unternehmen, durch Konzeptbildung und Vorgaben eine Rolle spielen und deshalb kollektiv für ihre Handlungen sowie deren absehbare Folgen Verantwortung tragen.[1160]

Datenverarbeitende sind für das Verarbeiten und Aufbereiten der Daten zuständig und können als kollektiver Akteur verstanden werden. In der Praxis umfasst dies viele verschiedene Aufgaben auf allen organisatorischen Ebenen eines Unternehmens. Die Entscheidung, wie transparent die Ergebnisse einer Datenverarbeitung sind, beeinflussen maßgeblich die Softwareentwickler – diese bekommen allerdings ihre Aufträge wiederum von der Managementebene.[1161] Solan-

1159 Friedman, Kahn 2008:1253; Friedman, Nissenbaum 1996:344
1160 French 1979:214; Isaacs 2016:9
1161 Gogoll et al. 2021:1089

ge die Entscheidungen zur Datenverarbeitung und -aufbereitung von einem kollektiven Akteur auf Grundlage von unternehmensinternen Entscheidungsstrukturen getroffen werden, trägt genau dieser kollektive Akteur die Verantwortung. Jedes betreffende Unternehmen sollte sich durch Reflexion der konkret beteiligten Mitarbeitenden über diese Verantwortung bewusst werden, damit jeder einzelne Schritt bedacht gegangen werden kann.

8.2.3 Zwischenfazit

Abschließend lässt sich also festhalten, dass Datengenerierende für den Schutz der Privatsphäre der Datenquellen verantwortlich sind, und Datenverarbeitende sind verantwortlich für die Ermöglichung der Transparenz für die Datennutzenden. Bei diesen kollektiven Akteuren kann es sich um verschiedene oder dasselbe Unternehmen handeln. Unabhängig davon gilt, dass die Verantwortlichen mit den Verantwortungsbereichen der jeweils anderen durchaus auch zu tun haben. Denn bei der Erhebung der Daten durch die Datengenerierenden muss zum Schutz der Privatsphäre immer schon feststehen, zu welchem Zweck diese Daten verarbeitet werden sollen. Deshalb haben auch die Datengenerierenden eine gewisse Kontrolle darüber, zu welchem Zweck die von ihnen erhobenen Daten verarbeitet werden. Dementsprechend tragen die Datengenerierenden nicht nur Verantwortung dafür, dass die Datenquellen über die Verarbeitungszwecke informiert werden, sondern auch dafür, dass die Zwecke vertretbar sind und dass die erhobenen Daten darauf zugeschnitten sind. Dabei haben Datengenerierende allerdings keine Kontrolle darüber, mit Hilfe welcher Algorithmen die Daten verarbeitet und wie die Ergebnisse aufbereitet werden – das liegt außerhalb ihres Aufgabenbereiches und damit auch außerhalb ihrer Verantwortung. Hier beginnt der Bereich der Datenverarbeitenden. Diese sind nun nicht nur für die Verarbeitung der Daten und der Aufbereitung der Ergebnisse verantwortlich, sondern insbesondere auch für die Qualität und Aussagekraft der Daten, die dabei verwendet werden. Denn während Datenverarbeitende nicht entscheiden, welche Daten

erhoben werden und wie die Datenquellen informiert werden[1162], kontrollieren sie doch, welche der erhobenen Daten bei der Verarbeitung der Daten verwendet werden und zum Beispiel beim Training des dafür verwendeten Algorithmus eine Rolle spielen. Diese Überprüfung der Daten und ihre Bereinigung von Verzerrungen ist essenziell, um für Datennutzende Transparenz herstellen zu können.

Im Folgenden soll betrachtet werden, wie genau es den Verantwortlichen möglich ist, ein Datenökosystem verantwortungsvoll zu gestalten.

1162 Angenommen, es handelt sich bei Datengenerierenden und Datenverarbeitenden um zwei getrennte kollektive Akteure.

9 Ethisches Gestalten

Bisher wurde in dieser Arbeit eine deontologische Risikoethik entwickelt, mit deren Hilfe die Zulässigkeit von Risikoauferlegung bewertet werden kann. Daraufhin wurden die beiden Risiken, die in Datenökosystemen entstehen, genau betrachtet, um herauszufinden, unter welchen Bedingungen diese Risiken entstehen und welche Pflichten daraus erwachsen. Daraus ließen sich administrative Anforderungen ableiten, welche erfüllt werden sollten, um vor den Risiken schützen zu können und die Auferlegung der Risiken ethisch zulässig zu machen. Es wurde außerdem klar, dass Datengenerierende und Datenverarbeitende verantwortlich sind für die ethische Gestaltung von Datenökosystemen, aber es stellt sich die Frage, wie sich die bisherigen ethischen Überlegungen in die Praxis umsetzen lassen. Angenommen ein Unternehmen möchte ein Datenökosystem entwickeln, wie kann es die ethischen Überlegungen aus dieser Arbeit in die Entwicklung einfließen lassen?

Da die Entwicklung eines Datenökosystems sehr umfangreich ist, wird diese im Folgenden in ihre verschiedenen Verarbeitungsschritte zerlegt betrachtet. Diese verschiedenen Schritte eines Datenökosystems sind das Sammeln, Speichern und Verarbeiten von Daten, sowie das Bereitstellen ihrer Ergebnisse. Jeder dieser Schritte bedarf ein Zusammenspiel zwischen sozialen Akteuren und der von ihnen genutzten Technologie. Die Technologie ist es, die es den menschlichen Akteuren ermöglicht, Daten zu sammeln, zu speichern, zu verarbeiten und Ergebnisse bereitzustellen. Auf Grund potenzieller Gefahren durch den Einsatz von datenverarbeitenden Technologien sowie auf Grund von deren Auswirkungen auf moralische Handlungen und Entscheidungen haben Technologien und ihre Verwendung eine moralische Signifikanz.[1163] Mit Hilfe der Ethik konnten in dieser Arbeit Antworten gefunden werden auf die Frage danach, unter welchen

1163 Kiran et al. 2015:10

Umständen es zulässig ist, diese verschiedenen Schritte durchzuführen.

So gibt die in dieser Arbeit entwickelte deontologische Risikoethik eine Antwort auf die Frage, wie Datenökosysteme gestaltet werden sollten – nämlich als gerechtes soziales System der Risikobereitschaft von dem alle profitieren und in dem niemand als bloßes Mittel gebraucht wird. Was genau das in der Praxis bedeutet, kann die Ethik allein nicht beantworten. Die KI-Ethikerin Anaïs Rességuier und die Rechtswissenschaftlerin Rowena Rodrigues stimmen zu: „The objective of ethics itself is not to impose particular behaviours and to ensure these are complied with."[1164] Die Umsetzung ethischer Überlegungen ist an empirische Gegebenheiten gebunden, welche die Ethik selbst nicht antizipieren kann. Deshalb reicht es nicht, allein eine ethische Antwort auf Gestaltungsfragen zu haben, sondern es bedarf einer Methode zur Umsetzung, damit diese ethische Antwort in der Praxis Anklang findet.

Um zu verstehen, in welchem organisatorischen Kontext eine solche Umsetzungsmethode zur Anwendung kommt, wird im Folgenden beleuchtet, wie technologische Gestaltungsprozesse in Organisationen funktionieren. Das heißt, es wird beleuchtet wie die Akteure mit Kontrolle über das Vorgehen im Datenökosystem – Datengenerierende und Datenverarbeitende – im Allgemeinen organisatorisch strukturiert sind. Der deutsche Ethiker und Anthropologe Arne Manzeschke definiert Organisationen als gesellschaftliche Intermediäre: „In ihnen wird ein Großteil der Entscheidungen verfertigt, die das Leben des Einzelnen betreffen und das Zusammenleben in der Gesellschaft strukturieren."[1165] In einer Organisation gibt es jedoch auch Entscheidungsprobleme, weil viele verschiedene Interessen in Einklang gebracht werden müssen. Manzeschke identifiziert hier mehrere Konfliktlinien: „Die Interessen und Einschätzungen der verschiedenen fachlichen Professionen, die fachlichen und ökonomischen Prioritäten und eng damit verkoppelt die technischen Möglichkeiten, das Interesse des Individuums auf Berücksichtigung als besonderer „Einzelfall" und das Interesse der Organisation, möglichst reibungsfrei die verschiedenen Fallgruppen nach einem Stan-

1164 Rességuier, Rodrigues 2020:2
1165 Manzeschke 2021:220

dard abzuarbeiten."[1166] Diese Konfliktlinien erschweren auch schon ohne ethische Überlegungen Entscheidungen in einer Organisation. Nach Ansicht des Soziologen Dirk Baecker erzeugt eine getroffene Entscheidung dann aber Gewissheit und Bestimmtheit, obwohl viele Unsicherheiten und Widersprüche bestehen.[1167]

Laut Manzeschke ist es Aufgabe der von ihm so genannten „Organisationsethik", mit Hilfe ethischer Reflexion Moral in Entscheidungsprozesse von Organisationen einzubinden.[1168] Ethik in einer Organisation habe die Aufgabe, die vielfältigen moralischen Auffassungen und die unterschiedlichen Arten der Entscheidungsfindung der Akteure zusammenzuführen und sie „(...) in einer komplexitätsadäquaten Theorie [zu] konzeptualisieren."[1169] Manzeschke bezeichnet dabei „Entscheidungen als Zielpunkt ethischer Bemühungen (...)".[1170] Weil ethische Fragen nie ohne Widerspruch abschließend geklärt werden können, kann eine organisatorische Entscheidung eine gewisse Bestimmtheit bieten.

An dieser Stelle stellt sich die Frage, auf welcher Organisationsebene die ethischen Fragen verhandelt werden sollten. Dem Verhaltensökonomen und Ethiker Jan Gogoll und seinen Kolleg:innen zufolge sind viele Probleme, die scheinbar auf die Software zurückzuführen sind, in Wirklichkeit das Ergebnis von Geschäftsmodellen oder politischen, rechtlichen und kulturellen Bedingungen. Deshalb sollten ethische Herausforderung nicht nur auf der Ebene der Entwickler:innen behandelt werden. Sondern es müsse schon im Bereich der Unternehmensethik die ethische Frage geklärt werden, ob ein bestimmtes technisches Produkt überhaupt entwickelt werden sollte.[1171]

Abbildung 4 veranschaulicht die verschiedenen Einflüsse und Ebenen der ethischen Entscheidungsfindung bei der Entwicklung einer Software. Laut Gogoll et al. ist jede Organisation eingebettet in ein Geflecht aus gesellschaftlichen Erwartungen, rechtlichen Anforderungen und kulturellen Normen. Sobald feststeht, dass ein ge-

1166 Manzeschke 2021:228
1167 Baecker 1999:146
1168 Manzeschke 2021:220
1169 Ebd.:228f.
1170 Ebd.:224
1171 Ebd.:1086

plantes Produkt legal ist und es einen gesellschaftlichen Konsens bezüglich des Produktes gibt, kann die Geschäftsführung eines Unternehmens als Strategisches Management beschließen, das Produkt herzustellen. Als Nächstes wird ein Projektteam innerhalb des Unternehmens über die genauen Spezifikationen des Produkts entscheiden – dies ist die Phase der Konzeptentwicklung. In dieser Phase werden die Anforderungen an das Produkt definiert. Dabei werden auch ethische Fragen besprochen und entschieden. Dann erst beginnt der Entwicklungsprozess. Hier nehmen die Teams von Entwickler:innen Einfluss auf die Gestaltung des Produkts, aber nur innerhalb der engen Grenzen der Spezifikationen, die in der Phase der Konzeptentwicklung festgelegt wurden: „Naturally, these parameters will never be completely determined. Therefore, the development team has some leeway in the development of the product."[1172] Selbst scheinbar harmlose Designentscheidungen können große Auswirkungen haben. So kommt es zum Beispiel vor, dass Familiennamen von Software als ungültig eingestuft werden, weil sie einen nicht vorgesehenen Bindestrich enthalten oder weil der Name nur aus einem Buchstaben besteht, aber mindestens zwei erwartet werden. Durch solche scheinbar unwichtigen Vorgaben definieren Programmierende, was als Name anerkannt wird und was nicht.

Gogoll et al. beschreiben diese verschiedenen Stufen der ethischen Entscheidungsfindung anhand der Entwicklung eines Pflegeroboters. Zuerst stellt die Gesellschaft in einem politischen Rahmen die Frage, ob ältere Menschen durch Roboter gepflegt werden sollten. Die Antwort auf die Frage wird in Form eines rechtlichen Rahmens oder in Form gesellschaftlicher und kultureller Bedingungen festgehalten. Als nächstes fragt ein konkretes Unternehmen im Bereich des Strategischen Managements, ob sie diejenigen sein wollen, die so einen Pflegeroboter bauen und mit welchem Geschäftsmodell. Eine positive Antwort resultiert in einem Projekt, das gestartet wird. Dann stellt das Projektteam im Rahmen der Konzeptionalisierung des Produktes die Frage danach, welche Fähigkeiten der Roboter haben sollte und welche ethischen Risiken bei der Gestaltung vermieden werden sollten und antwortet mit konkreten Anforderungen an das Produkt. Daraufhin stellt das Team von Entwickler:innen im Entwicklungsprozess die Frage, wie unter Berücksichtigung der

1172 Gogoll et al. 2021:1087

Anforderungen des Projektteams konkrete Funktionen implementiert werden können. Die schwierige Aufgabe der technischen Umsetzung aller gesellschaftlicher und ethischer Anforderungen fällt den Entwickler:innen zu. Außerdem müssen sie alle verbleibenden ethischen Fragen klären bzw. an das Management oder das Projektteam zurückspielen.[1173]

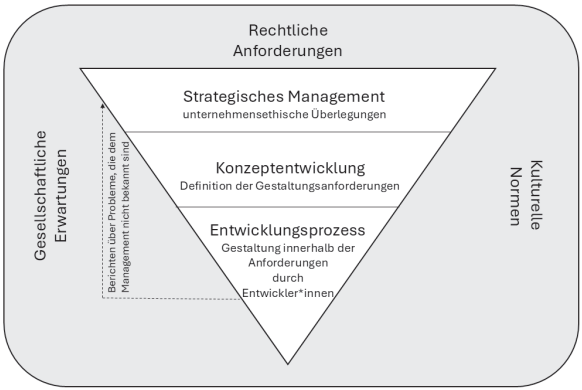

Abbildung 4 Entscheidungsebenen bei der Software-Entwicklung[1174]

Den medizinischen Informatiker:innen Maurice Mulvenna, Jennifer Boger und Raymond Bond zufolge ist es essentiell, dass alle Beteiligten die ethischen Aspekte einer Technologieentwicklung identifizieren, diskutieren und verstehen, da hier Funktionen implementiert werden können, welche für die Nutzenden der Technologie ethisch problematisch sind.[1175] Der Einsatz von ethischer Reflexion in einer Organisation kann aber laut Manzeschke auch Probleme zur Folge haben: „Wenn Organisationen sich auf eine Bearbeitung von moralischen Fragen einlassen, dann stehen sie in der Gefahr, die Kontrolle über die darin verhandelten Themen zu verlieren."[1176] So könnte zum Beispiel bei dem Entwicklungsprozess eines Pflegeroboters auf jeder der Entscheidungsebenen Uneinigkeit bestehen über morali-

1173 Gogoll et al. 2021:1091
1174 Leicht veränderte Darstellung nach Gogoll et al. 2021:1089
1175 Mulvenna et al. 2017:51; 53
1176 Manzeschke 2021:229

sche Fragen. Wenn es auf der Ebene des Strategischen Managements um die Frage geht, ob der Pflegeroboter gebaut werden sollte, dann halten das die Utilitarist:innen innerhalb des Managements wahrscheinlich für moralisch zulässig, weil dadurch das Pflegepersonal entlastet und die Patient:innen versorgt werden. Die Tugendethiker:innen innerhalb des Managements könnten hingegen argumentieren, dass die empathische Beziehung zwischen Pflegepersonal und Patient:in essenziell ist und es nicht zulässig wäre, diese durch einen Roboter ohne Einfühlungsvermögen zu ersetzen. Gleichermaßen könnten sich die Teammitglieder bei der Konzeptentwicklung uneinig sein, ob die Möglichkeit, dass der Pflegeroboter für lebendig gehalten wird, ein ethisches Risiko darstellt oder ob genau diese Täuschung eine hilfreiche Funktion des Roboters erfüllt. Selbst die Entwickler:innen im Entwicklungsprozess, die genaue Anforderungen erfüllen müssen, haben noch einen Spielraum, in welchem ethische Fragen zu Uneinigkeit führen können: „Developers can, for example, choose a technology that protects the user's privacy but still ensures the achievement of established business objectives."[1177] Fragen danach, ob über die Anforderungen aus der Konzeptentwicklung hinaus noch Schutzvorkehrungen getroffen werden sollten, können auch auf dieser Entscheidungsebene noch zu Konflikten führen.

Die Überforderung durch die Behandlung von moralischen Fragen stellt laut Manzeschke eine Bedrohung für die Handlungsfähigkeit der Organisation dar: „Moral, sofern sie in diesen Entscheidungsprozessen tatsächlich artikuliert wird, vervielfacht die Differenzen und Einschätzungen in Bezug auf eine Entscheidung, weshalb es eben auch starke Tendenzen gibt, der Moral bzw. den Moralen möglichst wenig Raum zu geben, um als Organisation handlungsfähig zu bleiben und nicht (zu) viel Zeit (=Geld) auf die Bearbeitung dieser Differenzen zu verwenden."[1178] Laut Manzeschke induziert Moral Differenzen und Krisen in Organisationen und wird deshalb, insbesondere in Unternehmen, aus Entscheidungsprozessen ausgeschlossen.[1179]

Ein weiterer Grund für den fehlenden Einsatz von Ethik kann dem KI-Ethiker Thilo Hagendorff zufolge ihre Wahrnehmung als

1177 Gogoll et al. 2021:1089
1178 Manzeschke 2021:228
1179 Ebd.:220

eine Behinderung für die Prozesse in der Organisation sein: „In business contexts, speed is everything in many cases and skipping ethical considerations is equivalent to the path of least resistance."[1180] Ethik wird oft nur als Hilfsmittel verstanden, das aber dem Gewinn des Unternehmens untergeordnet ist.[1181] Außerdem herrscht oft das Vorurteil, dass Ethik Aktivitäten stoppen und Handlungen verbieten oder vorschreiben will.[1182]

Genau dieses Behindern von Aktivität sollte jedoch laut Hagendorff nicht das Ziel der Ethik sein: „It should not be the objective of ethics to stifle activity, but to do the exact opposite, i.e. broadening the scope of action, uncovering blind spots, promoting autonomy and freedom, and fostering self-responsibility."[1183] Ethik kann im Gegenteil dazu beitragen, Aktivitäten zu fördern oder zu verbessern.[1184]

Es wird deutlich, dass innerhalb von Organisationen der Ethik gegenüber Skepsis herrscht. Um erfolgreiche Methoden der Technologieethik einzuführen, dürfen dem Technologieethiker Matt A. Murphy keine erheblichen Änderungen der bestehenden Prozesse gefordert werden: „They [the methods] must strike the balance of being rigorous while still remaining easy and relatively seamless to implement."[1185] Im Folgenden soll erörtert werden, wie sich diese schwierige Aufgabe in der Praxis umsetzen lässt. Da die Ethik allein keine praktischen Handlungsanweisungen bietet, soll nun eine Umsetzungsmethode gefunden werden, um die ethischen Erkenntnisse wirksam in die Praxis überführen zu können.

Es gibt verschiedenen Umsetzungsmethoden, die in der Literatur diskutiert und in der Praxis angewandt werden. Sie lassen sich grob in drei Kategorien einteilen: Ethische Leitlinien, die bei der Gestaltung der Technologie abgehakt werden können, einen aktiven Gestaltungsprozess, der durch ethische Werte geprägt ist oder Risikoanalysen, die vor der Entwicklung durchgeführt werden. Aus der bisherigen Arbeit lassen sich drei allgemeine Kriterien ableiten, die eine solche Umsetzungsmethode erfüllen sollte, damit sie bei der

1180 Hagendorff 2020:108
1181 Murphy 2023:4
1182 Boddington 2017:8
1183 Hagendorff 2020:112f.
1184 Boddington 2017:8
1185 Murphy 2023:4

Übersetzung meiner ethischen Erkenntnisse in die Praxis hilfreich sein kann:

Allgemeine Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung

i) Die Umsetzungsmethode sollte konkrete praktische Anleitung bieten.
ii) Die Umsetzungsmethode sollte den Einfluss der Technologie auf alle Betroffenen berücksichtigen.
iii) Im Rahmen der Umsetzungsmethode sollte die Verantwortung für die Umsetzung der ethischen Überlegungen bestimmten menschlichen Akteuren zugeordnet werden.

Dadurch, dass die Umsetzungsmethode konkrete praktische Anleitung bietet für die Ausrichtung technischer Entwicklungen an ethischen Maßstäben (i), wird es für eine Organisation möglich, ethische Überlegungen in die Praxis zu übersetzen. Außerdem bleibt die Orientierung an ethischen Überlegungen auf diese Weise kein Lippenbekenntnis, denn es fällt auf, wenn die konkrete praktische Anleitung nicht befolgt wird und somit bezüglich der ethischen Überlegungen keine Durchsetzungsmaßnahmen ergriffen werden. Damit der Einfluss der Technologie auf alle Betroffenen berücksichtigt werden kann (ii), muss der Vorteil aller betroffenen Personen das Ziel sein. Das heißt, die Risiken für eine Gruppe der betroffenen Personen können nicht einfach abgewogen werden gegen die Vorteile von anderen Personen – dann wäre die Umsetzungsmethode nicht geeignet für eine Technologieentwicklung an ethischen Maßstäben. Die Verantwortung für die Umsetzung der ethischen Überlegungen in die Praxis sollte bestimmten menschlichen Akteuren zugeordnet werden (iii), da auf diese Weise ein Akteur sowohl vorausschauend das Wohl aller Menschen im Blick hat als auch im Schadensfall für negative Folgen geradesteht.[1186] Bei den Verantwortungsträgern kann es sich allerdings nicht um Technologien handeln. In Kapitel 8 wurde deutlich, dass es für einen ethischen Umgang mit Technologien wichtig ist, dass Personen oder Kollektive von Personen die Verantwortung tragen für potenzielle Schäden im Nachhinein und für die

1186 Hart 1968:212; Young 2011:104; Danaher 2016:299; Kiener 2022:584; Nyholm 2024:192

durchdachte Gestaltung im Vorhinein – die Technologie kann diese Verantwortung nicht übernehmen.

Diese drei Kriterien können als notwendige Bedingungen für die ethische Gestaltung eines jeden technischen Teils eines Datenökosystems verstanden werden. Denn die drei allgemeinen Kriterien für die gesuchte Umsetzungsmethode müssen alle erfüllt sein, bevor überhaupt in Betracht gezogen werden kann, dass die betrachtete Methode zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben auch für gesamte Datenökosysteme geeignet ist. Dabei gelten dann zusätzlich noch drei Kriterien, die sich aus der bisherigen Arbeit ableiten lassen:

Spezifische Kriterien für die Umsetzungsmethode zur ethischen Gestaltung eines Datenökosystems

I. Die Umsetzungsmethode sollte es ermöglichen, die Auferlegung von Risiken zu regulieren.
II. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen zu erfüllen.
III. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen.

Bei der ethischen Gestaltung eines Datenökosystems ist es wichtig, die Risikoauferlegung zu problematisieren, welcher die Datenquellen durch Datengenerierenden und die Datennutzenden durch Datenverarbeitende ausgesetzt sind. Die Auferlegung von Risiken kann reguliert werden (I), wenn es möglich ist, das Risiko zu identifizieren und wenn klar ist, unter welchen Umständen das Risiko zulässig ist. Des Weiteren sollten die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen erfüllt werden (II), da die Auferlegung des Risikos einer verletzten Privatsphäre der Datenquellen nur so Teil eines gerechten sozialen Systems der Risikobereitschaft sein kann. Es muss sichergestellt werden, dass beim Sammeln der Daten das Einverständnis der Datenquellen eingeholt wird, sie dabei vor unfreiwilliger Einwilligung geschützt werden, indem sie über jeden Verwendungszweck der Daten informiert werden und ihnen die Möglichkeit gegeben wird, ihre Einwilligung zu verweigern. Zusätzlich müssen alle Daten vor Zweckentfremdung

geschützt werden. Damit für die Datennutzenden das Risiko der fehlenden Transparenz Teil eines gerechten sozialen Systems der Risikobereitschaft sein kann, müssen ebenfalls die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden erfüllt werden (III). Hier muss es darum gehen, algorithmische Ergebnisse im zum Verständnis notwendigen Kontext darzustellen, sodass klar wird, wie die Ergebnisse entstanden sind und welche Aussagen sie machen können.

Im Folgenden werden verschiedene Umsetzungsmethoden mit Hilfe der aus meiner Arbeit abgeleiteten Kriterien auf ihre Eignung als Grundlage zur ethischen Gestaltung eines Datenökosystems geprüft. Die Umsetzungsmethoden gliedern sich in die drei Kategorien Ethische Leitlinien, aktiver Gestaltungsprozess, und Risikoanalysen.

9.1 Ethische Leitlinien

Eine Liste von Anforderungen wird in der Regel in Form von Ethischen Leitlinien oder von einem Verhaltenskodex erstellt. Die Technologieethiker:innen Kelly Laas, Michael Davis, und Elisabeth Hildt definieren einen Ethikkodex folgendermaßen: „A code of ethics is an authoritative formulation of the (morally permissible) standards governing the conduct of members of a group, just because they are members of that group."[1187] Die Verbindlichkeit eines Ethikkodexes begründet sich immer in der Moral.[1188] Ein Ethikkodex hat laut Laas et al. mehrere Funktionen. So kann er besondere Verhaltensnormen dokumentieren, erklären oder festlegen, insbesondere wenn Traditionen oder Gewohnheiten nicht mehr ausreichen. Ein solcher Kodex hilft Personen, die in der Praxis neu sind, zu lernen, wie sie sich verhalten sollen und erinnert Personen mit viel Erfahrung an das, was sie sonst vergessen würden. Zudem kann im Fall eines Streits mit Hilfe eines Ethikkodexes geschlichtet werden. Außerdem kann ein Ethikkodex dazu beitragen, dass Außenstehende verstehen, was von den Mitgliedern der Gruppe erwartet werden kann. Schließlich

1187 Laas et al. 2022:2
1188 Ebd.

kann auf Grund von einem Ethikkodex auch eine rechtliche Haftung gerechtfertigt werden.[1189]

In der akademischen Literatur wurden in den letzten Jahren viele verschiedene Ethikkodizes und Leitlinien für die Nutzung von Daten oder KI aufgestellt.[1190] Im Jahr 2019 wurden in einer Studie 84 verschiedene ethische Richtlinien für KI allein in der akademischen normativen Ethikliteratur gefunden – Dokumente aus dem öffentlichen und privaten Sektor sind dabei noch nicht berücksichtigt.[1191] Aber auch Non-Profit-Organisationen, Unternehmen und politische Institutionen geben ihre eigenen digitalen Ethikrichtlinien heraus. SO schreibt der Sozialwissenschaftler und Historiker Markus Schmitz: „Viele Institutionen wie die EU, die OECD und das Future of Life Institute sowie einige größere Technologieunternehmen haben bereits ethische Grundsätze zum Einsatz von KI und zur Nutzung von Daten veröffentlicht."[1192] Als Orientierungshilfe für die ethische Gestaltung der Datennutzung werden ethische Grundsätze genannt, die berücksichtigt werden sollten, damit alle von den Technologien profitieren und gleichzeitig die Risiken minimiert werden können.[1193] In den akademischen Leitlinien zur KI sind die wiederkehrenden Grundsätze: "(…) transparency, justice and fairness, nonmaleficence, responsibility and privacy."[1194]

Als erstes fällt auf, dass die Herkunft dieser Grundsätze unklar ist. Laas et al. argumentieren, dass Ethische Leitlinien abhängig davon, wer sie verfasst, unterschiedlich ausfallen, da jede Gruppe von Autor:innen ihre eigene Sicht der Dinge hat. Für die Fachleute in dem betreffenden Bereich können ganz andere Aspekte einer neuen Technologie wichtig sein als für die Öffentlichkeit oder für Ethiker:innen.[1195] Um den Einfluss der Technologie auf alle Betroffenen zu berücksichtigen (ii), erscheint es sinnvoll, alle Perspektiven in den Ethischen Leitlinien und Kodizes zu repräsentieren. Allerdings gibt es im Kontext dieser Umsetzungsmethode keine Anleitung dafür, wie entschieden wird, welche Grundsätze eine Rolle spielen sollen.

1189 Laas et al. 2022:2
1190 Laas et al. 2022:3; Jobin et al. 2019:391
1191 Jobin et al. 2019:391
1192 Schmitz 2020:368
1193 Ebd.
1194 Jobin et al. 2019:391
1195 Laas et al. 2022:4f.

Die Grundsätze sind in allen Sektoren ähnlich, aber die Häufigkeit einzelner Grundsätze variiert je nachdem, wer die Leitlinien verfasst hast. So haben die KI-Forscher Yi Zeng, Enmeng Lu und Cunqing Huangfu herausgefunden, dass der Grundsatz „Privatsphäre" in den von Regierungen erstellten Leitlinien sehr präsent ist, in denen von Wissenschaftler:innen dagegen kommt er statistisch gesehen deutlich seltener vor und im Unternehmenssektor sogar noch seltener.[1196] Aber selbst wenn die gleichen Grundsätze hervorgehoben werden, unterscheiden sich ihre Auslegung und die Ideen für ihre Umsetzung erheblich.[1197] Der Gedanke findet sich auch bei dem KI-Experten Sebastian Hallensleben und seinen Kolleg:innen: „Values that form the basis for AI ethics are open to different interpretations and may vary depending on an AI system's application context."[1198] Abhängig vom Kontext, in dem ein KI-System eingesetzt wird, unterscheiden sich die Ansprüche an seine ethischen Werte. So ist zum Beispiel der Anspruch an Privatsphäre und Fairness höher im Justizbereich als in der industriellen Produktion.[1199]

Durch die unterschiedlichen Ansprüche an die ethischen Werte in verschiedenen Kontexten besteht in der praktischen Umsetzung Unsicherheit darüber, welche ethischen Grundsätze Vorrang haben und wie Konflikte zwischen ethischen Grundsätzen gelöst werden sollten.[1200] Ein solcher Konflikt könnte zum Beispiel dann entstehen, wenn durch das Sammeln möglichst vieler personenbezogener Daten einerseits die algorithmischen Ergebnisse weniger verzerrt und diskriminierend würden, aber andererseits die einzelnen Personen durch das Sammeln der Daten in ihrer Privatsphäre beschränkt würden.[1201] Es ist nicht klar geregelt, ob das Entdiskriminieren der Ergebnisse oder der Privatsphäreschutz der Datenquellen Vorrang hat und wie dieser Konflikt in der Praxis aufgelöst werden soll. Ethische Prinzipien werden in der Literatur für den Kontext der Praxis als nutzlos, ineffektiv und zahnlos bezeichnet.[1202] Dem Informatiker

1196 Zeng et al. 2019:3, Figure 3; Gogoll et al. 2021:1094
1197 Jobin et al. 2019:391
1198 Hallensleben et al. 2020:10
1199 Ebd.
1200 Jobin et al. 2019:396
1201 Jobin et al. 2019:396
1202 Rességuier, Rodrigues 2020:1; Munn 2023:869, 871; Hagendorff 2020:99; Mulvenna et al. 2017:51

Luke Munn zufolge ist es deshalb leicht die ethischen Prinzipien zu ignorieren: „The failure of AI ethical principles is not spectacular but silent, resulting in the desired outcome: business as usual."[1203] Dass trotz der Anerkennung ethischer Prinzipien diese in der Praxis ignoriert werden deutet darauf hin, dass zwischen den Prinzipien und ihrer praktischen Umsetzung eine Diskrepanz besteht. Selbst wenn man sich auf Prinzipien geeinigt hat, bewirkt das noch keine tatsächliche Veränderung in der Gestaltung algorithmischer Systeme.[1204]

Die Softwareentwickler Andrew McNamara, Justin Smith and Emerson Murphy-Hill fanden in einer Studie mit sowohl Studierenden der Softwaretechnik als auch professionellen Softwareentwickler:innen heraus, dass die ausdrückliche Anweisung, einen ethischen Kodex zu berücksichtigen, keinen erkennbaren Unterschied in der Arbeit gemacht hat – die Studienteilnehmer:innen haben ihre bewährten Arbeitsmethoden nicht auf Grund des ethischen Kodex geändert.[1205] Die bloße Einführung von Ethikgrundsätzen in einer Organisation ändert also das Verhalten der Entwickler:innen nicht.[1206]

Es gibt Gründe für diese beobachtbare Ineffizienz der ethischen Prinzipien. Einer der Gründe ist, dass ein Abweichen von einem Ethik-Kodex keine Konsequenzen nach sich zieht und es auch sonst keinen Mechanismus für die Durchsetzung von Leitlinien gibt.[1207] Dieses Problem könnte laut Schmitz zum Beispiel durch den Einsatz eines Ethikbeirates gelöst werden, welcher die Ethischen Leitlinien des betreffenden Unternehmens ausarbeitet und dann ihre Einhaltung unterstützt und kontrolliert.[1208]

Laut Gogoll et al. ist ein weiterer Grund für die fehlende Umsetzung ethischer Prinzipien in die Praxis die Unterbestimmtheit („underdetermination") von ethischen Werten. Sie fungierten als allgemeine Aussagen, die alleine keine konkrete oder praktische Orientierung bieten können.[1209] Mittelstadt stimmt zu: „AI Ethics initiatives have thus far largely produced vague, high-level principles and value statements which promise to be action-guiding, but in practice

1203 Munn 2023:872
1204 Morley et al. 2020:2147
1205 McNamara et al. 2018:732
1206 Murphy 2023:3
1207 Hagendorff 2020:99, 113; Schmitz 2020:369f.; Munn 2023:871
1208 Schmitz 2020:371
1209 Gogoll et al. 2021:1095

provide few specific recommendations (…)."[1210] Ohne ausreichende Konkretisierung und Kontextualisierung sind Software-Ingenieure mit der Aufgabe, moralische Werte zu beachten, auf sich allein gestellt. Eine Schwierigkeit daran ist, dass Werte nicht prozessorientiert sind und somit auch die Mittel, mit denen sie erreicht werden können, nicht logisch erschlossen werden können. Deshalb werden laut Gogoll et al. Werte oft erst im Nachhinein betrachtet, anstatt das Handeln von vorneherein danach auszurichten. Hinzu kommt, dass es sich Menschen gerne leicht machen und am liebsten automatische Lösungen finden durch Anwendung von Dispositionen, Konventionen oder moralischen Regeln – das ist aber bei Widersprüchen nicht mehr zielführend.[1211] Das größte Problem bei Ethischen Leitlinien ist also das Fehlen konkreter praktischer Anleitung (Kriterium i).

Ein weiterer Grund dafür, dass ethische Prinzipien nicht in die Praxis umgesetzt werden, könnte sein, dass manche Unternehmen kein ethisches Fundament haben. Denn ohne eine breitere moralische und ethische Grundlage können Prinzipien in der Praxis missachtet werden.[1212] Nur weil ein Unternehmen ethische Richtlinien und Checklisten einführt, bedeutet das noch nicht, dass dadurch ethische Produkte entwickelt werden – im Gegenteil: Unternehmen können den Anschein von Ethik genießen, ohne dass diese Substanz hat.[1213] So können mit Hilfe der Einführung ethischer Prinzipien kritische Stimmen beruhigt werden, während sich aber an den kritisierten Praktiken nichts ändert.[1214]

Diese Beibehaltung der bestehenden Praxis ist manchmal genau das, was durch das Aufstellen von ethischen Prinzipien erzielt werden soll, denn dadurch wird laut Floridi der Regulierung durch Gesetze zuvorgekommen[1215]: „(…) ethics may be exploited as if it were an alternative to legislation (…)"[1216] Ethische Leitlinien werden manchmal als nichtlegislative politische Instrumente oder auch „Soft Law" bezeichnet.[1217] Diese ethischen Prinzipien sind aber kein Er-

1210 Mittelstadt 2019:1
1211 Gogoll et al. 2021:1098
1212 Lauer 2021:24
1213 Munn 2023:872
1214 Hagendorff 2020:100
1215 Rességuier, Rodrigues 2020:2; Floridi 2019:188
1216 Floridi 2019:188
1217 Jobin et al. 2019:389

satz für gesetzliche Regulierung, da sie keine konkreten Verhaltensregeln vorschreibt[1218] und, wie bereits erläutert, oft nur schwer in die Praxis übersetzbar ist. Die Ethischen Leitlinien sollten als Methode zur Umsetzung der Ethik bei dieser Übersetzung behilflich sein. Doch der Interpretationsspielraum, den Ethische Leitlinien zulassen, ermöglicht die Ausnutzung der Ethik durch zum Beispiel Ethics Shoppping und Ethics Washing.[1219]

Mit Ethics Shopping bezeichnet man die Maßnahme, bestimmte ethische Grundsätze aus einer Vielzahl möglicher Grundsätze auszuwählen, um bereits bestehende Verhaltensweisen im Nachhinein zu rechtfertigen.[1220] Gogoll et al. stellen fest, dass keine kohärente Ethik verwirklicht werden kann, wenn Verhaltenskodizes als Ansammlungen von Werten verstanden werden, aus denen wir mehr oder weniger willkürlich Werte auswählen können.[1221] Weil dieses Ethics Shopping möglich ist, sind die Autor:innen der Meinung, dass Ethische Leitlinien und Verhaltenskodizes keine normative Orientierung bieten. Diese Orientierungsfunktion werde außerdem durch eine Indifferenz verhindert. Denn ein bestimmter Verhaltenskodex könne zur Rechtfertigung unterschiedlicher oder sogar widersprüchlicher Handlungen herangezogen werden.[1222]

Ethics Washing bezeichnet ein Vorgehen, bei dem man nur den Anschein erweckt, sich ethisch zu verhalten, aber tatsächlich Fehlinformationen über ethische Werte verbreitet oder oberflächliche Maßnahmen zugunsten dieser Werte ergreift.[1223] So können zum Beispiel ethische Prinzipien wie Gerechtigkeit passend zu den Produktmerkmalen definiert werden, sodass man behaupten kann, man hielte ethische Grundsätze ein, ohne sich aber auf die nötige Reflexion oder Umgestaltung einzulassen.[1224]

Es stellt sich also die Frage, wie ethische Prinzipien so konkretisiert werden können, dass ihre Umsetzung ermöglicht wird und dass gleichzeitig Ethics Shopping und Ethics Washing verhindert werden

1218 Rességuier, Rodrigues 2020:2
1219 Ebd.; Floridi 2019:188
1220 Floridi 2019:186
1221 Gogoll et al. 2021:1097
1222 Ebd.:1097
1223 Floridi 2019:187
1224 Munn 2023:870

können. Murphy weist auf folgenden Lösungsansatz hin: „One popular way to try and concretize ethical principles is to create ethical checklists to be filled out alongside development."[1225] Den Technologiephilosoph:innen Asle H. Kiran, Nelly Oudshoorn und Peter-Paul Verbeek zufolge vernachlässigt eine statische Checkliste allerdings, dass ethische Grundsätze durch neue technische Entwicklungen beeinflusst werden können und auch dass in verschiedenen kulturellen Umfeldern unterschiedliche ethische Ansprüche herrschen.[1226] Alternativ können laut Murphy ethische Prinzipien auch konkretisiert werden, indem diese Prinzipien als ethische Fragen formuliert werden.[1227] Hier erkennt allerdings Murphy selbst ein Problem: „The more open-ended or unstructured questions are in methods like this, the more they invite unstructured and unsystematic answers."[1228] Hagendorff ist der Meinung, dass Ethische Leitlinien wirksamer sein könnten, wenn sie durch detailliertere technische Erläuterungen ergänzt würden: „The fact that not a single prominent ethical guideline goes into greater technical detail shows how deep the gap is between concrete contexts of research, development, and application on the one side, and ethical thinking on the other."[1229] Die Ergänzung der Ethischen Leitlinien um technische Details erfordert jedoch eine Übersetzung der Theorie in die Praxis.

Das VCIO (Values, Criteria, Indicators, Observables) Modell ist eine Methode zur Spezifizierung und Operationalisierung von ethischen Werten, die von Hallensleben et al. entwickelt wurde. Diese Methode soll ethische Grundsätze praktikabel, vergleichbar und messbar machen.[1230] Denn ethische Werte sollten menschliches Handeln leiten. Sie enthalten jedoch keine Standards zur Bewertung ihrer Umsetzung. Deshalb sollen nach dem VCIO-Modell andere Komponenten diese Aufgaben erfüllen – und zwar Kriterien, Indikatoren und beobachtbare Größen. Die Kriterien legen fest, wann Werte erfüllt sind oder verletzt werden. Um wiederum feststellen zu können, ob ein Kriterium erfüllt ist, werden Indikatoren benötigt.

1225 Murphy 2023:2
1226 Kiran et al. 2015:6
1227 Murphy 2023:3
1228 Ebd.:3
1229 Hagendorff 2020:111
1230 Hallensleben et al. 2020:8

Diese prüfen, ob die Kriterien erfüllt sind. Schließlich muss es noch beobachtbare Größen geben. Diese quantifizieren oder qualifizieren inwieweit die Indikatoren erfüllt sind. Die Kriterien bilden auf einer abstrakten Ebene notwendige Bedingungen für die Erfüllung von Werten und die Indikatoren stellen notwendige Bedingungen für die Erfüllung der Kriterien dar. Sie werden in deliberativen Prozessen aufgestellt und festgelegt.[1231]

Das sogenannte „AI Ethics Label" soll der Übersichtlichkeit dienen. Dieses Label ist inspiriert von der Energieverbrauchskennzeichnung. Mit Hilfe von Buchstaben und Farben soll auf einen Blick verständlich sein, inwiefern ein bestimmter Wert erfüllt wird. Dafür werden die verschiedenen beobachtbaren Größen für einen bestimmten Wert in einer Bewertung zusammengefasst. Wenn zum Beispiel der Wert „Transparenz" mit Hilfe des „AI Ethics Labels" dargestellt werden soll, dann werden die zur Erfüllung notwendigen Bedingungen in Form von Kriterien und ihren Indikatoren betrachtet. Die beobachtbaren Größen quantifizieren oder qualifizieren, inwieweit die Indikatoren erfüllt sind und können dann mit A bis G bewertet werden. Der Durchschnitt dieser Bewertungen wird dann auf dem Label angezeigt.[1232]

Bei der Bewertung eines Algorithmus ist ein Kriterium des Wertes „Transparenz" die „Zugänglichkeit". Indikatoren für dieses Kriterium sind dann die Frage danach, wer Zugang zu dem Algorithmus hat und ob es für den Konfliktfall eine Schlichtungsstelle gibt. Beobachtbare Größen sind dann die Antworten auf diese Fragen. So könnte jede Person gleichermaßen Zugang zu dem Algorithmus haben – das wäre mit einer guten Bewertung (A) zu versehen – oder es hat niemand Zugang zu dem Algorithmus – das wäre mit einer schlechten Bewertung (G) zu versehen. Ebenfalls ist beobachtbar, ob es für den Konfliktfall eine (kompetente) Schlichtungsstelle gibt (A) oder nicht (G). Aus dem Durchschnitt der Bewertungen ergibt sich dann die Gesamtbewertung, welche anzeigt, inwiefern der zu bewertende Algorithmus das Kriterium „Zugänglichkeit" des Wertes „Transparenz" erfüllt.[1233]

1231 Hallensleben et al. 2020:15ff.
1232 Ebd.:31f.
1233 Ebd.:21

Die Anforderungen an einen Algorithmus sind unterschiedlich, je nachdem in welchem Kontext er eingesetzt wird. Deshalb haben die Sozioinformatiker:innen Tobias D. Krafft und Katharina Zweig eine Risikomatrix aufgestellt, welche das Risiko für den potenziellen Schaden eines KI-Systems in Abhängigkeit von seiner Nutzung aufzeigt. Dabei wird einerseits die Intensität des potenziellen Schadens eines KI-Systems bewertet, in dem die Auswirkungen auf individuelle Rechte und die Gesellschaft als Ganzes betrachtet werden. Andererseits wird die Abhängigkeit der potenziell betroffenen Parteien von der algorithmischen Entscheidung bewertet, in dem die Möglichkeiten zur Vermeidung des potenziellen Schadens betrachtet werden.[1234]

Krafft und Zweig leiten fünf verschiedene Regulierungsklassen ab, die unterschiedlich hohe Anforderungen an Transparenz und Nachvollziehbarkeit stellen. Während für KI-Systeme in Klasse 0 – bei zum Beispiel der personalisierten Auswahl von Werbung für Kleidung – keine Transparenzpflichten gefordert werden, sollen KI-Systeme in Klasse 4 – bei zum Beispiel der personalisierten Auswahl von Werbung für medizinische Dienstleistungen – nur dann eingesetzt werden, wenn dem Risiko nachweisbar ein hoher Gesamtnutzen entgegensteht. In Klasse 1 bis 3 steigen die Anforderungen an Transparenzpflichten mit dem Risiko, welches das KI-System mit sich bringt.[1235]

Mit dem VCIO-Modell gelingt es, ethische Prinzipien so zu konkretisieren, dass ethischen Leitlinien oder Kodexe konkrete praktische Anleitung bieten können und somit Kriterium i) erfüllen. Auch Kriterium iii) wird erfüllt, da die Verantwortung für die ethische Gestaltung der betreffenden Technologie im VCIO-Modell bestimmten menschlichen Akteuren zugeordnet wird.[1236] Insbesondere durch die Risikomatrix von Krafft und Zweig Betrachtung wird deutlich, dass die Auswirkungen der Technologie auf alle potenziell betroffenen Parteien berücksichtigt wird und also auch Kriterium ii) erfüllt ist.

Auch wenn die ethischen Prinzipien mit Hilfe des VCIO-Modells konkretisiert werden und praktische Anleitung bieten können, bleibt

1234 Krafft, Zweig 2019:19ff.; Hallensleben et al. 2020:37f.
1235 Hallensleben et al. 2020:38ff.; Krafft, Zweig 2019:31
1236 Hallensleben et al. 2020:27

die Herkunft dieser Prinzipien weiter unklar. Denn Hallensleben et al. schreiben: „For the proposed AI Ethics Label, we carefully suggest six values (…) based on contemporary discourse and operability."[1237] Im Kontext des VCIO-Modells wurden die ethischen Werte, die spezifiziert werden sollten, auf der Grundlage des heutigen Diskurses und ihrer Operationalisierbarkeit ausgewählt. Es handelt sich bei den sechs Werten nur um einen Vorschlag, aber trotzdem birgt die fehlende Begründung der Auswahl der Werte die Gefahr von Ethics Shopping. Die Relevanz im zeitgenössischen Diskurs ist keine gute Begründung, denn der Diskurs kann manche Werte vernachlässigen und kann außerdem seine Schwerpunkte verändern. Zudem sind Werte nicht deshalb relevant, weil sie operationalisierbar sind – diese Perspektive ist pragmatisch, aber erfüllt nicht den ursprünglichen Zweck der Umsetzung ethischer Werte in die Praxis.

Im VCIO-Modell ist die Wahl der ethischen Werte nicht gut begründet, aber die in dieser Arbeit betonten ethischen Werte wurden aus der epistemischen Asymmetrie, Kontrollungleichheit und dadurch ermöglichten Risikoauferlegung im Datenökosystem hergeleitet. Mit Hilfe der Umsetzungsmethode der Ethischen Leitlinien könnte also die von mir begründeten ethischen Prinzipien der Privatsphäre und Transparenz spezifiziert werden. Da sich die administrativen Anforderungen zur Umsetzung von Privatsphäre und Transparenz mit Sicherheit in den Indikatoren für Privatsphäre und Transparenz wiederfinden würden, wären dadurch auch Kriterium (II) und (III) erfüllt. Auf diese Weise ließe sich zumindest ein Teil meiner ethischen Überlegungen in die Praxis übersetzen. Aber Kriterium (I) erfüllt die Umsetzungsmethode der Ethischen Leilinien nicht. Denn sie ermöglicht es nicht, die Risikoauferlegung zu problematisieren und kann dadurch auch nicht die Auferlegung von Risiken in der Praxis regulieren. Da aber das Kriterium (I) als notwendige Bedingung für die ethische Gestaltung eines Datenökosystems verstanden werden kann, eignen sich Ethische Leitlinien nicht zur Umsetzung meiner ethischen Überlegungen in die Praxis.

1237 Hallensleben et al. 2020:7

9.2 Gestaltungsprozess

Anstatt der Anwendung Ethischer Leilinien bei der Entwicklung von Technologien, gibt es auch die Möglichkeit, ethische Überlegungen direkt in den Gestaltungsprozesses bei der Technologieentwicklung zu integrieren. Dadurch werden die ethischen Überlegungen lebendig gehalten, anstatt sie in feste Leitlinien zu gießen. Das ist laut Rességuier, Rodrigues notwendig: „AI ethics is in a deep and vital need of the continuously renewed process of questioning the world and the lenses through which we see it – consciously, constantly and iteratively."[1238] Diese KI- oder Technologieethik kann als ein reflektierender Prozess zur Entwicklung möglichst ethisch gestalteter Technologien verstanden werden, der unter anderem dabei hilft, Voreingenommenheit unter den am Gestaltungsprozess Beteiligten aufzudecken.[1239] Die Technologiephilosoph:innen und Digitalethiker:innen Jessica Morley, Luciano Floridi und Libby Kinsey sowie die Psychologin Anat Elhalal sind der Meinung, dass ein solcher Wechsel von ethischen Grundsätzen hin zu einem ethischen Gestaltungsprozess die Ethik in der Praxis nachvollziehbarer macht.[1240]

Der österreichische Designer Victor Papanek definiert einen Gestaltungsprozess als „Das Planen und Konzipieren von Handlungen auf ein erwünschtes, absehbares Ziel hin (...)."[1241] Ein Prozess kann also gestaltet werden, indem man sich bewusst um eine sinnvolle Ordnung bemüht.[1242] Ob eine solche Ordnung sinnvoll oder richtig ist, hängt laut Papanek davon ab, welche Bedeutung wir ihr zumessen.[1243] Er schreibt: „Design muss sinnvoll sein."[1244] Aber das Wort „sinnvoll" könne durch alle möglichen Begriffe ersetzt werden, welche dann den Zweck des Designs, und damit seine Funktion, definieren. Kann die Funktion eines Designs zum Beispiel sein, schön oder abstrakt zu wirken, so kann sie auch sein, ethisch durchdacht zu sein.

1238 Rességuier, Rodrigues 2020:4
1239 Morley et al. 2021:246f.
1240 Ebd.:252
1241 Papanek 2009:20
1242 Ebd.:20
1243 Papanek 2009:21
1244 Ebd.:22

Der erste Schritt in einem Gestaltungsprozess ist das Formulieren von Anforderungen an den zu gestaltenden Gegenstand. Dabei spielen viele Akteure eine Rolle – Personen, die ein Interesse an dem zu gestaltenden Gegenstand haben oder Personen, die von seinen Auswirkungen betroffen sein könnten, wie zum Beispiel Auftraggebende und Kund:innen.[1245] Die Anforderungen werden nicht einmalig zu Beginn definiert, sondern sie werden während des gesamten Gestaltungsprozesses weiter formuliert.[1246] In einem Gestaltungsprozess wird nicht nur die optimale Lösung für bestimmte Ziele und Anforderungen gesucht, sondern es werden auch Entscheidungen getroffen, die über die ursprünglich formulierten Anforderungen und Ziele hinausgehen.[1247]

Manchmal entsteht aber auch das Problem, dass nicht alle Anforderungen an die Gestaltung gleichzeitig erfüllt werden können. Eine mögliche Lösung im Engineering Bereich ist das Formulieren von Schwellenwerten für bestimmte Anforderungen, bei deren Überschreitung der Entwurf als inakzeptabel angesehen wird – diese Methode nennt sich „satisficing". Die Sicherheitsethikerin Anke van Gorp und der Technologieethiker Ibo van de Poel definieren sie folgendermaßen: „Satisficing means that a designer is satisfied with any solution that is "good enough" given the design requirements and will not look for optimal solutions."[1248] Dieser Ansatz minimiert die Komplexität der Gestaltung, sie führt jedoch nicht zu den ethisch wünschenswertesten Lösungen. Ethisch problematisch wird es, wenn zum Beispiel Sicherheitsanforderungen überdacht oder verwässert werden, weil sie mit anderen Anforderungen kollidieren.[1249]

Hier besteht laut der Technologieforscherin Barbara Ribeiro und ihren Kolleg:innen ein „(...) issue of societal alignment (...)"[1250] – eine Problematik der Anpassung an die Gesellschaft. Diese beschreibt die Notwendigkeit, eine bessere Abstimmung zwischen den Zielen von Innovationen und den Zielen der Öffentlichkeit zu erreichen. Das heißt, bei der Gestaltung von Innovationen soll sicherge-

1245 Van Gorp, van de Poel 2001:17
1246 Ebd.:18
1247 Ebd.:18
1248 Ebd.:19
1249 Ebd.:21
1250 Ribeiro et al. 2018:318

stellt werden, dass die Innovationsergebnisse mit den Werten und Bedürfnissen der Öffentlichkeit in Einklang gebracht werden. Die Herausforderung besteht dabei darin, die Öffentlichkeit einzubeziehen, gesellschaftliche Bedürfnisse zu formulieren und die Ziele der Innovationen auf die Erfüllung dieser Bedürfnisse abzustimmen.[1251]

Gestaltungsprozessmethoden zur Umsetzung von ethischen Überlegungen in die Praxis sind in der Literatur in sieben verschiedene Ausgestaltungen prävalent: Embedded Values, Embedded Ethics, Ethics by Design, Human-centered Design, Privacy by Design, Responsible Innovation und Value-Sensitive Design. Im Folgenden werden die ersten vier nur kurz vorgestellt, da sie Kriterium i) nicht erfüllen und keine konkrete praktische Anleitung bieten. Sie begründen aber, warum die ethische Abwägung bezüglich der praktischen Anwendung von Technologien zur Datenverarbeitung schon bei ihrer Gestaltung essentiell ist. Die letzten drei Methoden werden etwas ausführlicher dargestellt, um zeigen zu können, dass keine von ihnen alle Kriterien erfüllt.

9.2.1 Vier Ansätze

Embedded Values

Der Embedded Values Ansatz geht von dem Verständnis aus, dass die Werte der Designer in die Technologie einfließen. In Bezug auf eine ethische Produktentwicklung ist der Anspruch, dass die gestalteten Technologien wünschenswerte Werte fördern oder unterstützen und dadurch wünschenswerte Praktiken aufrechterhalten.[1252] In der Literatur gibt es drei verschiedene Positionen dazu, wie Werte in technologische Entwürfe einfließen[1253]: *verkörpert*, *exogen* und *interaktionell*. Die *verkörperte Position* besagt, dass die Designer

1251 Ribeiro et al. 2018:318, 324

1252 Zuber et al. 2022:3–5

1253 Dem Technikforscher Patrick Feng zufolge haben in den 1970er Jahren die Sozialwissenschaftler*innen Technologie noch als wertneutral verstanden. Im Jahr 2000 jedoch schreibt er, dass die Wissenschaftler*innen sich mittlerweile einig sind, dass Technologien Werte verkörpern, die sich in ihrer Gestaltung widerspiegeln (Feng 2000:212).

ihre eigenen Absichten und Werte in die Technologie übertragen.[1254] Die *exogene Position* geht davon aus, dass gesellschaftliche Kräfte maßgeblich beeinflussen, wie eine Technologie genutzt wird. Technologische Systeme seien deshalb keinesfalls wertneutral, sondern begünstigen die Interessen von Menschen mit wirtschaftlicher und politischer Macht.[1255] Die *interaktionale Position* besagt, dass die Merkmale oder Eigenschaften, die Menschen in Technologien einfließen lassen, bestimmte Werte eher unterstützen und andere behindern, während die tatsächliche Nutzung der Technologie von den Zielen der Menschen abhängt, die mit ihr interagieren.[1256]

Der Embedded Values Ansatz macht die Designer als bestimmten kollektiven Akteur dafür verantwortlich, dass sie ihre ethischen Werte im Rahmen des Gestaltungsprozesses in die betreffende Technologie einfließen lassen (iii). Das widerspricht im Prinzip nicht dem Ziel, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt werden soll (ii), aber da es sich bei den Designern um einen sehr kleinen Teil der Gesellschaft handelt, besteht die Gefahr, dass sie Nachteile für andere gesellschaftliche Gruppen übersehen. Außerdem fehlt die konkrete praktische Anleitung aus Kriterium i), da unklar ist, welche Schritte man gehen müsste, um ethische Werte in die Technologie einzubetten. Da schon die allgemeinen Kriterien nicht vollständig erfüllt sich, erübrigt es sich die spezifischen Kriterien zu betrachten. Somit eignet sich der Embedded Values Ansatz nicht zur Umsetzung von ethischen Überlegungen in die Praxis.

Embedded Ethics

Der Ansatz der Embedded Ethics hat laut dem Medizinethiker Stuart McLennan und seinen Kolleg:innen ein übergeordnetes Ziel: „The overarching objective of an embedded ethics approach is to help develop AI technologies that are ethically and socially responsible, and that benefit and do not harm individuals and communities."[1257] Es gibt allerdings unterschiedliche Wege, auf denen dieses Ziel erreicht werden kann. So wird der Begriff Embedded Ethics ver-

1254 Friedman, Kahn 2008:1242f.
1255 Ebd.:1243
1256 Ebd.:1243f.
1257 McLennan et al. 2020:488

wendet für die Verknüpfung von Ethik und automatisiertem Denken in autonomen Maschinen[1258], aber auch für integrierte Unterrichtseinheiten zum Thema ethisches Denken im Informatik-Lehrplan[1259] und schließlich für das Integrieren einer Ethikerin oder eines Ethikers in den Entwicklungsprozess einer Technologie[1260].

Der McLennan et al. vertreten die Ansicht, dass durch die integrierten Ethiker:innen erklärt und begründet werden sollte, warum der verwendete ethische Ansatz geeignet ist, um die spezifischen Ziele des Projekts zu erreichen.[1261] Die Verantwortung tragen also die einzelnen integrierten Ethiker:innen (iii). Da es ihre Aufgabe ist, die Entwicklung von Technologien zu begleiten, die dem Einzelnen und der Gemeinschaft nutzen und nicht schaden[1262], ist das Kriterium, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird, erfüllt (ii). Jedoch ist auch bei dem Ansatz der Embedded Ethics das Fehlen der konkreten praktischen Anleitung für seine Umsetzung ein Problem (i). Denn wie genau Maschinen programmiert werden müssten, um ethische Empfehlungen geben zu können, was genau Informatikstudierende oder -auszubildende im Unterricht lernen müssten, und was genau die Aufgaben der integrierten Ethiker:innen sind, über das Begründen des verwendeten ethischen Ansatzes hinaus, bleibt unklar. Weil Kriterium i) nicht erfüllt ist, wird auch der Embedded Ethics Ansatz in dieser Arbeit nicht weiterverfolgt.

Ethics by Design

Der Ethics by Design Ansatz umfasst die proaktive Identifizierung und Behandlung potenzieller ethischer Fragen während des Designprozesses.[1263] Bei der Entwicklung von Technologien, die deutliche Auswirkungen auf unsere Gesellschaft haben könnten, stellt sich die Frage, wie sie mit den Werten in Einklang gebracht werden können, auf denen die Menschenrechte und das Wohlergehen beru-

1258 Bonnemains et al. 2018
1259 Grosz et al. 2019
1260 McLennan et al. 2020:488; McLennan et al. 2022:3
1261 Ebd.:5
1262 Ebd.:488
1263 Kellmeyer 2024:251

hen.[1264] Zur Beantwortung dieser Frage gibt es aber dem auf Technologierecht spezialisierten Lachlan Urquhart und dem Technologieforscher Peter J. Craigon zufolge nicht nur einen Ansatz: „(…) there are many approaches to doing 'ethics by design', often using different labels (…)."[1265] Es gibt nicht nur verschiedene Ansätze zur ethischen Gestaltung von Technologien mit vielen verschiedenen Namen, sondern es werden sogar die gleichen Bezeichnungen für unterschiedliche Herangehensweisen verwendet. So bedeutet Ethics by Design laut der Informatikerin Virginia Dignum und ihren Kolleg:innen, dass Maschinen so programmiert werden können, dass sie sich nach ethischen Grundsätzen verhalten.[1266] Laut Mulvenna et al. erfordert der Ethics by Design Ansatz jedoch eine Zusammenarbeit zwischen Technik und Ethik während des gesamten Gestaltungsprozesses.[1267] Deshalb schlagen sie vor, dass eine ständige internationale Arbeitsgruppe eingerichtet wird, die auf der Basis von 12 Prinzipien Grundsätze ableitet.[1268]

Ich halte es für gefährlich, Kriterium iii) nicht zu erfüllen, stattdessen die Verantwortung der Technik zu übertragen und zu behaupten, dass es möglich ist, Ethik einzuprogrammieren.[1269] Da eine Ethik komplexer ist als ein heruntergebrochener Verhaltenskodex, würde diese Programmierung oft einer ethischen Entscheidung nicht gerecht. Wenn man sich trotzdem darauf verlässt, können unethische Entscheidungen unbemerkt getroffen werden, welche die Nutzenden womöglich stark beeinflussen.

Leider bleibt bei Mulvenna et al. unklar, woher die 12 Prinzipien stammen und wie sie sich in einem Gestaltungsprozess umsetzen lassen, sodass hier Kriterium i) nicht erfüllt ist. Da jeweils mindestens eins der allgemeinen Kriterien nicht erfüllt wird, fehlt auch dem Ethics by Design Ansatz etwas Wesentliches bei der Umsetzung von ethischen Überlegungen in die Praxis.

1264 Dignum et al. 2018:64
1265 Urquhart, Craigon 2021:96
1266 Dignum et al. 2018:65
1267 Mulvenna et al. 2017:51, 53
1268 Ebd.:53f.
1269 Dignum et al. 2018:65

Human-centered Design

Human-centered Design bedeutet, dass der Mensch nicht nur als Teil des Systems betrachtet wird, sondern in jedem Aspekt des Designs eine zentrale Rolle spielt. Dabei fungiert der Designer als Übersetzer von gesellschaftlichen Bedürfnissen in das Design von KI-Systemen.[1270] Der KI-Ethiker Philipp Kellmeyer ist davon überzeugt, dass durch eine Einbeziehung aller Stakeholder in den Entwicklungsprozess (Stakeholder-Befragung) die daraus resultierenden Technologien technisch solide sind und wirklich auf die Bedürfnisse und Werte derer eingehen, denen sie helfen sollen.[1271] Die Umsetzung dieser Einbindung der Stakeholder in den Gestaltungsprozess erfordere jedoch eine engagierte Gemeinschaft von Stakeholdern, eine inklusive Integration aller Stakeholder und den Ausbau ihrer Fähigkeiten durch iteratives Lernen.[1272]

Der Human-centered Design Ansatz hat die Funktion, auf die Rolle der Menschen im Gestaltungsprozess aufmerksam zu machen. Dadurch sollen ethische Anforderungen an zu gestaltende Technologien identifiziert werden. Das Problem bei der Methode der Stakeholder-Befragung ist jedoch ihre fehlende Skalierbarkeit. So können kontextspezifische Lösungen gefunden werden für die Gruppe der befragten Stakeholder, jedoch lassen sich diese Lösungen nur schwer auf größere Populationen übertragen.[1273] Es ist zwar das Ziel des Human-centered Designs, nur Technologien zu entwickeln, die den Einfluss der Technologie auf alle Betroffenen berücksichtigen (ii) und somit zum Vorteil für alle von der Technologie betroffenen Personen sind, es ist aber unklar, ob dieses Ziel mit Hilfe der Stakeholder-Befragung erreicht werden kann. Außerdem ist die Identifikation ethischer Anforderungen durch die Stakeholder Beteiligung zwar ein wichtiger erster Schritt zur ethischen Gestaltung von Technologien, aber hier fehlen konkrete Umsetzungsvorschläge (i). Und es bleibt unklar, wer die Verantwortung für die Umsetzung eines der Human-centered Design Ansätze tragen soll (iii). Da die allgemeinen Kriterien nicht (mit Sicherheit) erfüllt werden, reicht auch

1270 Auernhammer 2020:1320
1271 Kellmeyer 2024:252
1272 Ebd.:254
1273 Ebd.:257f.

der Human-centered Design Ansatz nicht aus zur Umsetzung von ethischen Überlegungen in die Praxis.

9.2.2 Privacy by Design

Ein konkreter Umsetzungsvorschlag findet sich in der Privacy by Design (PbD) Methode. Hier wird die Privatsphäre als ethische Anforderung an technische Systeme identifiziert. Die Methode stammt von Ann Cavoukian, der Informations- und Datenschutzbeauftragten der kanadischen Stadt Ontario: „Privacy by Design (PbD), the framework to proactively embed privacy directly into information technology, business practices, physical design, and networked infrastructures – making it the default."[1274] Bei der Privacy by Design Methode sollen Datenschutzprinzipien proaktiv in ein System integriert werden, durch zum Beispiel Standardeinstellungen zum Schutz der Privatsphäre sowie Sicherheitsmaßnahmen. Zudem soll durch die Methode der Schutz der Privatsphäre in den Risikomanagementprozessen der Unternehmen anerkannt werden.[1275]

Beim Schutz der Privatsphäre geht es darum, dass so wenige personenbezogene Daten wie möglich erhoben werden und der Einzelne die größtmögliche Kontrolle über seine personenbezogenen Daten hat. Laut Cavoukian muss eine Organisation alle Aspekte des Schutzes der Privatsphäre und der Informationssicherheit unter Kontrolle haben, um von sich behaupten zu können, dass sie die Fähigkeit hat, Rechenschaft ablegen zu können.[1276] Idealerweise sollten technische Systeme laut der Wirtschaftsinformatikerin Sarah Spiekermann so aufgebaut werden, dass die Kontrolle der Nutzenden maximiert und die Beteiligung von Dienstanbietern minimiert wird. Wenn dieses Ideal nicht umsetzbar ist, sollte der Schutz der Privatsphäre unterstützt werden durch technisch durchsetzbare Standardrichtlinien („Opt-out"-Einstellungen) oder Datenknappheitsrichtlinien, Datenübertragbarkeit und Benutzerzugriffs- und Löschrechte.[1277]

1274 Cavoukian 2012:18
1275 Spiekermann 2012:38
1276 Cavoukian et al. 2010:409
1277 Spiekermann 2012:40

Die Ziele des „Privacy by Design" können durch die Einhaltung von sieben Grundprinzipien erreicht werden: Die Technologie soll erstens proaktiv und präventiv anstatt reaktiv gestaltet werden, sodass Verletzungen der Privatsphäre von vorneherein verhindert werden können. Zweitens gilt der Schutz der Privatsphäre als Standard, das heißt, die Nutzenden müssen nicht aktiv ihre Privatsphäre schützen, denn der Schutz der Privatsphäre ist drittens bereits in das Design eingebettet. Dabei wird aber viertens die Funktionalität des Systems nicht beeinträchtigt, da andere Geschäftsinteressen mit dem Schutz der Privatsphäre koexistieren. Fünftens sind starke Sicherheitsmaßnahmen als Grundlage für den Schutz der Privatsphäre unerlässlich. Des Weiteren sind sechstens Sichtbarkeit und Transparenz unerlässlich, um Verantwortlichkeit und Vertrauen zu schaffen. Das siebte Grundprinzip von Privacy by Design ist der Respekt für die Privatsphäre der Nutzenden, denn die Gestaltung von Technologie sollte bewusst auf die Interessen und Bedürfnisse der einzelnen Nutzenden abgestimmt sein, sodass diese ihre eigenen personenbezogenen Daten verwalten können.[1278]

Es gibt laut Cavoukian unendlich viele Möglichkeiten, wie Unternehmen den Schutz der Privatsphäre kreativ in ihre Abläufe und Produkte integrieren können[1279]: „(…) the task for privacy-aware engineers and systems architects is to translate the PbD conceptual framework into a set of specific, and operationally feasible, tools."[1280] Die bereits erwähnten Marie Caroline Oetzel und Sarah Spiekermann haben sich dieser Aufgabe angenommen und das Privacy Impact Assessment (PIA) entwickelt. Der PIA-Rahmen schlägt konkrete Ziele für den Schutz der Privatsphäre vor und beschreibt sieben Schritte, um diese zu erreichen.[1281] Bei der Ausführung von PIAs werden Bedrohungen der Privatsphäre untersucht und dazu angeregt, schon in einem frühen Stadium des Gestaltungsprozesses Schutzmaßnahmen einzuplanen und diese fortlaufend zu überprüfen. So können Probleme beim Schutz der Privatsphäre während der

1278 Cavoukian et al. 2010:409f.
1279 Ebd.:413
1280 Cavoukian 2012:19
1281 Spiekermann 2012:39

Produktentwicklung erkannt und proaktiv Maßnahmen zur Verbesserung des Schutzes in Systeme eingebaut werden.[1282]

Dank der konkreten Grundprinzipien von Cavoukian und des Privacy Impact Assessments von Oetzel und Spiekermann, bietet die Privacy by Design Methode eine praktische Anleitung zur Umsetzung des Wertes der Privatsphäre in die Gestaltung von Technologien (i). Der Einfluss der Technologie auf jede Person, deren Privatsphäre durch die Technologie betroffen ist, wird berücksichtigt (ii) und die Zuschreibung von Verantwortung (iii) für die einzelnen Schritte fällt dem die PIA durchführenden Unternehmen zu.[1283] Es wird also durch die Privacy by Design Methode nicht klar, wem die Verantwortung für die Gestaltung zugeordnet wird, sondern nur wer sie zuordnen soll. Selbst wenn man das durchgehen ließe, und damit alle allgemeinen Kriterien erfüllt wären, scheitert die Privacy by Design Methode an mindestens einem der spezifischen Kriterien für die ethische Gestaltung eines Datenökosystems. Denn die Privacy by Design Methode betrachtet zwar ausführlich die Umsetzung der Privatsphäre (II), jedoch spielen andere im Datenökosystem relevante Werte wie die Transparenz keine Rolle. Deshalb ermöglicht es die Privacy by Design Methode nicht, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen, sodass Kriterium III nicht erfüllt werden kann. Aus dem Grund wird auch die Privacy by Design Methode nicht weiterverfolgt als Lösung zur Umsetzung von ethischen Überlegungen in die Praxis.

9.2.3 Responsible Innovation

Ein weiterer konkreter Umsetzungsvorschlag dazu, wie sich technische Entwicklungen an ethischen Maßstäben ausrichten lassen, stellt die Responsible Innovation Methode dar (auch als „Responsible Research and Innovation" bezeichnet). Ursprünglich bezieht sich der Begriff der Innovation auf Neuerungen aller Art.[1284] Mittlerweile

1282 Oetzel, Spiekermann 2014:126ff.
1283 Ebd.:142
1284 Van den Hoven 2013:80

wird der Innovationsbegriff jedoch den bereits erwähnten Technologiephilosophen Vincent Blok und Pieter Lemmens zufolge nur noch bei der Erforschung neuer Technologien und der kommerziellen Nutzung dieser neuen Technologien eingesetzt.[1285] Bei verantwortungsvoller Innovation (Responsible Innovation) geht es darum, soziale und ethische Aspekte der Erforschung und Nutzung von Technologien ausdrücklich zu berücksichtigen und sie mit wirtschaftlichen, soziokulturellen und ökologischen Aspekten in ein Gleichgewicht zu bringen.[1286] Die Berücksichtigung von ethischen und gesellschaftlichen Bedenken soll dadurch umgesetzt werden, dass die von dem Innovationsprozess direkt Betroffenen (Stakeholder) und Vertreter der Öffentlichkeit in den Innovationsprozess integriert werden.

Verantwortung wird im Rahmen der Responsible Innovation Methode als proaktiv und nicht als reaktiv betrachtet, das heißt, es geht nicht um Rechenschaftspflicht für potenziell unerwünschte Ergebnisse, sondern um aktive Gestaltung von Innovationsprozessen für gesellschaftlich wünschenswerte Ziele.[1287] Laut dem bereits erwähnten Technologieforscher Owen et al. ist Responsible Innovation philosophisch verankert in den vorausschauenden Dimensionen der Verantwortung: Sorgfalt und Reaktionsfähigkeit (care und responsiveness). Diese Dimensionen betonen den Technologieforscher:innen Mirjam Burget, Emanuele Bardone und Margus Pedaste zufolge, dass es nicht um die Ergebnisse geht, sondern um den Prozess selber: „(…) *taking care* describes more an attitude rather than a result."[1288] Reaktionsfähigkeit soll bedeuten, dass Reflexion und Deliberation mit Handlungen verknüpft werden, die einen wesentlichen Einfluss auf die Innovation haben.[1289]

Owen et al. definieren im Jahr 2013 Responsible Innovation folgendermaßen: „Responsible innovation is a collective commitment of care for the future through responsive stewardship of science and innovation in the present."[1290] Diese kollektive Verpflichtung

1285 Blok, Lemmens 2015:28
1286 Jakobsen et al. 2019:2339; Blok, Lemmens 2015:20
1287 Stahl et al. 2021:177
1288 Burget et al. 2017:14
1289 Owen et al. 2013:29
1290 Ebd.:36

zur Sorge für die Zukunft lässt sich bei der Entwicklung neuer Technologien umsetzen, indem der Innovationsprozess antizipativ, reflexiv, deliberativ und responsiv gestaltet wird.[1291] Der Gestaltungsprozess ist antizipativ, wenn die Folgen des zu gestaltenden Produktes analysiert werden. Er ist reflexiv, wenn ethische Implikationen des Produktes analysiert werden. Der Gestaltungsprozess ist deliberativ, wenn die Perspektiven, Visionen, Fragen und Probleme der Stakeholder in Bezug auf das Produkt berücksichtigt werden. Und der Prozess ist responsiv, wenn ethische Probleme zu Veränderungen im Prozess und im Produkt führen.[1292] Dadurch, dass der verantwortungsvolle Innovationsprozess und insbesondere die Eigenschaft der Deliberation inklusiv ist, wird zumindest im Prinzip allen Menschen die Teilhabe ermöglicht.[1293]

Um wirksam zu sein, darf es sich bei den vier Kernanforderungen (Antizipation, Reflexivität, Deliberation und Responsivität) nicht nur um Grundsätze handeln, sie müssen institutionell eingebettet sein in die Innovation und in den Innovationsprozess.[1294] Eine konkrete Umsetzung kann jedoch flexibel an den Anwendungskontext angepasst werden. So könnte zum Beispiel das Team, das den Innovationsprozess voranbringt, zum verantwortungsvollen Innovieren gebracht werden, in dem in Entscheidungsprozessen immer wieder die passenden Fragen gestellt werden: Woran arbeiten Sie gerade? (Reflexiv) Warum arbeiten Sie daran? (Reflexiv, Deliberativ) Wie könnten Sie es anders angehen? (Responsiv) Wer könnte in Zukunft betroffen sein? (Antizipativ).[1295]

Die Antworten des Teams könnten dann wiederum von einem unabhängigen Gremium bewertet werden anhand von fünf Kriterien, mit denen die Kernanforderungen der Responsible Innovation geprüft werden. Erstens müssen die Hauptrisiken ermittelt, verwaltet und für akzeptabel befunden werden (Reflexiv), zweitens muss das Vorgehen den einschlägigen Vorschriften entsprechen (Reflexiv), drittens muss allen relevanten Parteien klar mitgeteilt werden, welchen Zweck die Innovation erfüllt, um die Stakeholder zu informie-

1291 Owen et al. 2013:44
1292 Niehaves, Heger 2015:172; Owen et al. 2013:38
1293 Burget et al. 2017:14
1294 Owen et al. 2013:39
1295 Ebd.:42

ren und eine ausgewogene Diskussion zu fördern (Reflexiv, Deliberativ), viertens müssen künftige Anwendungen und Auswirkungen beschrieben und bei neuen Informationen überprüft werden (Antizipativ, Reflexiv), fünftens müssen Mechanismen ermittelt werden, um die Ansichten der Öffentlichkeit und der Interessengruppen zu den möglichen Anwendungen und Auswirkungen zu verstehen (Deliberativ, Reflexiv).[1296]

Die Wirtschaftswissenschaftler Rob Lubberink, Johan van Ophem und Onno Omta sowie der Technologiephilosoph Vincent Blok entwickeln aus einer Literaturanalyse eine andere Strategie zur Operationalisierung von Responsible Innovation. Sie übersetzen die vier Kernanforderungen (zuzüglich Inklusion und Wissensmanagement) in eine Liste von wichtigen Aktivitäten und dazugehörigen Strategien.[1297] Die Kernanforderung der Antizipation umfasst dabei das frühzeitige Erkennen erwünschter und unerwünschter Auswirkungen von Innovationen durch die Beobachtung gesellschaftlicher und ökologischer Bedürfnisse, um daraus konkrete Innovationsanforderungen abzuleiten.[1298] Reflexivität bedeutet, dass Unternehmen kritisch über ihre eigenen Werte, Motive, Handlungen und Wissenslücken nachdenken und deren Einfluss auf den Innovationsprozess analysieren.[1299] Deliberation wird durch den offenen, strukturierten Austausch zwischen Stakeholdern auf Basis von genauen und transparenten Informationen unter Berücksichtigung der Interessen aller Stakeholder erreicht.[1300] Reaktionsfähigkeit wird erreicht, indem Organisationen durch den Abbau bürokratischer Hürden und der Neuausrichtung interner Rollen die Innovationsprozesse flexibel an veränderte Bedingungen und Bedürfnisse anpassen.[1301]

Lubberink et al. fügen den vier Kernanforderungen der Responsible Innovation außerdem die Anforderungen der Inklusion und des Wissensmanagements hinzu: Inklusion bezieht sich auf die Einbindung vielfältiger Stakeholder in den Innovationsprozess, wobei eine faire Verteilung von Aufgaben und Nutzen sowie die Stärkung von

1296 Stilgoe et al. 2013:1575; Owen et al. 2013:43
1297 Lubberink et al. 2017:11–18
1298 Ebd.:11
1299 Ebd.:13
1300 Ebd.:15
1301 Lubberink et al. 2017:16f.

Beziehungen im Fokus stehen.[1302] Wissensmanagement umfasst die gezielte Gewinnung und Integration relevanten Wissens in den Innovationsprozess, um Innovationen im Einklang mit den Bedürfnissen der Stakeholder zu entwickeln – dies wird unterstützt durch Austauschplattformen und flache Hierarchien.[1303] Lubberink et al. argumentieren, dass das Übersetzen der Kernanforderungen von Responsible Innovation in die wichtigsten Aktivitäten und die dazugehörigen Strategien dabei helfen kann, alle von der Innovation betroffenen Akteure einzubeziehen und dadurch den Innovationsprozess so zu steuern, dass letztlich nachhaltige, gesellschaftlich wünschenswerte und ethisch akzeptable Lösungen erzielt werden.[1304]

Es ist deutlich erkennbar, dass die Responsible Innovation Methode konkrete praktische Anleitung bietet (i), weil auf Grundlage ihrer Kernanforderungen bestimmte Leitfragen und Umsetzungsstrategien entwickelt wurden. Das Ziel des Berücksichtigens des Einflusses der Technologie auf alle Betroffenen soll durch das Einbeziehen der Stakeholder erreicht werden (ii) – hier gibt es jedoch Umsetzungsschwierigkeiten in der Praxis. Die Betriebswirtschaftlerinnen Luciana Maines da Silva und Claudia Cristina Bitencourt stellen zusammen mit Kadígia Faccin und Tatiana Iakovleva aus dem Innovationsmanagement bei einer Literaturanalyse von empirischen Studien zur Umsetzung von Responsible Innovation fest, dass die Stakeholder häufig erst einbezogen werden, wenn die Innovation bereits auf dem Markt ist.[1305] Auch dem Technologiephilosophen und -ethiker Bernd Carsten Stahl und seinen Kolleg:innen fällt beim Erproben der Umsetzung dieser Methode auf, dass Responsible Innovation trotz der ursprünglichen Intention, Stakeholder von Anfang an einzubeziehen, in der Praxis oft erst dann eingesetzt wird, wenn die wichtigsten Entscheidungen bereits getroffen sind.[1306] Problematisch ist das, weil auf diese Weise das Feedback von Stakeholdern nur eingeschränkt berücksichtigt werden kann und womöglich nicht alle

1302 Lubberink et al. 2017:14
1303 Ebd.:17f.
1304 Ebd.:23
1305 Silva et al. 2019:16
1306 Stahl et al. 2021:189

Risiken antizipiert werden können.[1307] Die Umsetzung bleibt also hinter dem Anspruch zurück bei der Entwicklung der Technologie ihren Einfluss auf alle Betroffenen zu berücksichtigen und dadurch einen Vorteil für alle von der Technologie betroffenen Personen zu bewirken (ii). Insbesondere auch, weil es in der Praxis schwierig ist nachzuvollziehen, ob die Deliberation mit den Stakeholdern direkte Auswirkungen auf die Innovation hat und ob die Ergebnisse des Innovationsprozesses den Wünschen der Stakeholder entsprechen.[1308]

Außerdem ist auch die Zuordnung der Verantwortung bei der Umsetzung der Responsible Innovation Methode unklar (iii). Denn laut Lubberink et al. führt die Beratung mit den Stakeholdern zu einer kollektiven Verantwortung.[1309] Ich denke jedoch, Stakeholder sollten nicht verantwortlich gemacht werden, solange nicht nachvollziehbar ist, ob ihre Deliberation wirklich Auswirkungen hat auf die daraufhin entwickelte Technologie.[1310] Owen et al. schlagen hingegen eine Co-Verantwortung vor: Bei der Umsetzung von Responsible Innovation sollen nicht nur die Innovator:innen eine wichtige Rolle spielen, sondern auch die Institutionen, in denen Innovationen gestaltet werden, sowie diejenigen, die Innovation fördern. Insbesondere die Fördernden haben laut Owen et al. eine einzigartige Position, um die Responsible Innvoation institutionell zu verankern.[1311] So fällt die Verantwortung zur Umsetzung der Responsible Innovation allen am Entwicklungsprozess Beteiligten zu – das lässt aber auch befürchten, dass niemand sich konkret verantwortlich fühlt. Aus meiner Sicht müsste klar sein, welcher bestimmte menschliche Akteur dafür zuständig ist, die vier Kernanforderungen institutionell einzubetten, damit Kriterium iii) erfüllt wäre. Da die allgemeinen Kriterien ii) und iii) nicht erfüllt werden, erübrigt es sich die spezifischen Kriterien zu betrachten. So weist auch die Responsible Innovation Methode Mängel auf bei der Umsetzung ethischer Überlegungen in die Praxis.

1307 Silva et al. 2019:16
1308 Stahl et al. 2021:180f.
1309 Lubberink et al. 2017:2
1310 Stahl et al. 2021:180f.
1311 Owen et al. 2013:46

9.2.4 Value-Sensitive Design

Die Informationswissenschaftlerin Batya Friedman hat als erste die Methode des Value-Sensitive Design (VSD) vorgebracht und meint damit, dass man sich beim Gestalten und Implementieren von Computer Technologien auf menschliche Werte fokussieren sollte. Sie ruft schon 1997 dazu auf, dass Value-Sensitive Design als Teil der Kultur der Informatik begriffen werden soll und dass es als Kriterium dienen sollte, nach dem Systeme als mangelhaft und Designer als nachlässig beurteilt werden können.[1312] Werte verstehen sie und ihre Kollegen im Jahr 2008 als etwas, das eine Person oder eine Gruppe von Menschen im Leben für wichtig hält. Somit hängen Werte wesentlich von den Interessen und Wünschen der Menschen je nach kulturellem Milieu ab.[1313]

Bei dieser Umsetzungsmethode orientiert sich der Gestaltungsprozess an als wichtig und erstrebenswert empfundenen Werten mit Hilfe einer integrativen und iterativen, dreiteiligen Methodik, die aus *konzeptionellen*, *empirischen* und *technischen* Untersuchungen besteht. So wird *konzeptionell* untersucht, welche Stakeholder es gibt und welche Werte wichtig sein könnten. Dabei werden sorgfältige Überlegungen darüber angestellt, wie sich die eigenen technologischen Entwürfe auf die Stakeholder auswirken könnten. Die als wichtig identifizierten Werte müssen dann sorgfältig konzeptualisiert werden, da ein Begriff sehr unterschiedlich verstanden werden kann. Eine gut ausgearbeitete Definition der betreffenden Werte klärt dann schon die grundlegenden Fragen, die das jeweilige Projekt aufwirft, und bietet eine Grundlage für den Vergleich der Ergebnisse verschiedener Forschungsteams. *Empirisch* wird untersucht, welche Werte die Stakeholder für wichtig halten und durch welche Gestaltungsentscheidungen welche Werte realisiert werden können: Friedman beschreibt das mit dem bereits erwähnten Psychologen Peter H. Kahn Jr. und dem Informatiker Alan Borning so: „Empirical investigations encompass any human activity that can be observed, measured, or documented."[1314] Auf diese Weise kann auch der Erfolg und die Nutzbarkeit einer bestimmten Gestaltung bewertet

1312 Friedman 1997:12f.
1313 Friedman et al. 2008:70f.
1314 Friedman et al. 2002:3

werden. *Technisch* wird überprüft, wie bestimmte Eigenschaften von Technologien menschliche Werte unterstützen oder behindern, und es werden technische Lösungen gesucht, um die Werte umsetzen zu können. Die Annahme dabei ist, dass sich bestimmte Technologien für die Umsetzung bestimmter Werte eignen – diese Eignung folgt aus den Eigenschaften der Technologien.[1315]

Die drei Teile der Methodik sind voneinander abhängig und beeinflussen sich gegenseitig. Es gibt keine vorgeschriebene Reihenfolge für die drei Teile. Jedoch halten es die Wirtschaftsinformatiker:innen Till Winkler und Sarah Spiekermann für ratsam, erst konzeptionell zu arbeiten und dann empirisch zu analysieren, bevor technisch untersucht wird – insbesondere wenn für Wiederholungen des Prozesses die Zeit fehlt.[1316]

Value-Sensitive Design befasst sich in erster Linie mit Werten, die sich auf menschliches Wohlbefinden, Würde, Gerechtigkeit, Wohlfahrt und Rechte beziehen. Solche Werte haben laut dieser Methode einen moralisch-epistemischen Stellenwert, unabhängig davon, ob eine bestimmte Person oder Gruppe diese Werte vertritt – ihre Ausgestaltung kann aber in unterschiedlichen Kulturen oder zu verschiedenen Zeitpunkten variieren.[1317] Friedman und Kahn haben die folgenden zwölf Aspekte entwickelt, welchen sie als spezifische menschliche Werte mit ethischer Bedeutung einen ausgeprägten Anspruch auf Ressourcen im Designprozess zuordnen: Menschliches Wohlergehen, Eigentum und Besitz, Privatsphäre, Unvoreingenommenheit, universelle Nutzbarkeit, Vertrauen, Autonomie, informierte Zustimmung, Verantwortlichkeit, Identität, Gelassenheit und ökologische Nachhaltigkeit.[1318]

Laut Friedman und Kahn sind die ersten neun Aspekte traditionell und beruhen auf deontologischen und konsequentialistischen moralischen Orientierungen. Problematisch ist, dass diese beiden Orientierungen ganz unterschiedliche Werte hervorbringen – in Entscheidungssituationen beziehen sie sich sogar oft auf widersprüchliche Werte. Außerdem gibt es in keinem der beiden einen

1315 Friedman et al. 2008:71–73; Friedman et al. 2002:2f.; Friedman, Kahn 2008:1251; Niehaves, Heger 2015:173f.; Winkler, Spiekermann 2018:17
1316 Winkler, Spiekermann 2018:19
1317 Friedman, Kahn 2008:1250f.
1318 Ebd.:1251

abgeschlossenen Wertekanon, sodass sehr viele Werte als traditionell bezeichnet werden können. Die letzten drei Werte seien nicht traditionell innerhalb der Gemeinschaft von Designern und Forschern, die sich mit der Interaktion von Menschen und Computern auseinandersetzen.[1319] Hier besteht das Problem, dass es innerhalb der Gemeinschaft von Designern und Forschern, die sich mit der Interaktion von Menschen und Computern auseinandersetzen, noch viel mehr Werte gibt, die als nicht traditionell gelten können. Außerdem behaupte ich, dass informierte Zustimmung gar kein Wert ist – es handelt sich vielmehr um eine Maßnahme, die Werte wie Privatsphäre oder Autonomie schützen kann.

Wie schon bei den Ethischen Leitlinien, könnten die von mir aus der epistemischen Asymmetrie eines Datenökosystems hergeleiteten Prinzipien Privatsphäre und Transparenz als Werte gelten, an denen sich ein ethischer Gestaltungsprozess ausrichten sollte. Das heißt, die konzeptionelle Untersuchung der wichtigen Werte könnte sich auf die in dieser Arbeit identifizierten Konzepte stützen. Denn dank der dreiteiligen Methodik bietet die Value-Sensitive Design Methode konkrete praktische Anleitung (i) und es wird der Einfluss der zu entwickelnden Technologie auf alle Stakeholder sowohl konzeptionell als auch empirisch berücksichtigt (ii). Die empirische Berücksichtigung der Werte der Stakeholder bereitet der Value-Sensitive Design Methode jedoch ähnliche Schwierigkeiten, wie der Responsible Innovation Methode. So haben Winkler und Spiekermann bei der Analyse von 17 Studien, welche die Anwendung der Value-Sensitive Design Methode abbilden, festgestellt, dass oft kein guter methodischer Ansatz zur Identifizierung der Stakeholder, sondern nur grob definierte Stakeholder-Gruppen verwendet wurden. Hierdurch können Stakeholdern und ihre Werte unabsichtlich ausgeschlossen werden.[1320] Das gefährdet die Erfüllung von Kriterium ii).

Außerdem ist auch Kriterium iii) nicht eindeutig erfüllt, denn die Angaben dazu, wer die Verantwortung für die Umsetzung der ethischen Überlegungen tragen soll, variieren. Einerseits sollen Programmierende durch Value-Sensitive Design Sozialinformatik prak-

1319 Friedman, Kahn 2008:1251
1320 Winkler, Spiekermann 2018:19

tizieren und anwenden können[1321], andererseits haben in den Anwendungsbeispielen scheinbar Forscher:innen die Value-Sensitive Design Methode durchgeführt[1322]. An wieder anderer Stelle wird Fachleuten für Mensch-Computer-Interaktion (human-computer interaction professionals) auf Grund ihrer beruflichen Stellung eine besondere ethische Verantwortung zugeschrieben.[1323] Da also die allgemeinen Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung vom Value-Sensitive Design nicht alle erfüllt werden, erübrigt es sich die spezifischen Kriterien zur ethischen Gestaltung eines Datenökosystems zu betrachten. Deshalb wird auch die Value-Sensitive Design Methode im Folgenden nicht genutzt, um die Umsetzung von ethischen Überlegungen in die Praxis anzuleiten.

9.3 Risikoanalyse

Eine dritte mögliche Kategorie von Umsetzungsmethoden zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben ist die Betrachtung der Risiken, die durch den Einsatz der Technik entstehen können. Der Informatiker Peter G. Neumann meint, dass technische Risiken durch langfristige Planung vermieden werden können. Es sollten also nicht Risiken für hinreichend unwahrscheinlich erklärt werden oder nur Versicherungsschutz angestrebt werden, sondern es sollte proaktiv gehandelt werden, indem die Systeme mit Bedacht entwickelt werden.[1324] Im Folgenden werden drei verschiedene Methoden diskutiert, welche auf einer Risikoanalyse basieren: die Folgenabschätzung, das interne Audit und die ethische Risikoanalyse.

9.3.1 Folgenabschätzung

Der bereits erwähnte Technologierechtler Lachlan Urquhart und der Technologieforscher Peter J. Craigon haben ein praktisches

1321 Friedman et al. 2002:7
1322 Ebd.:4f.
1323 Friedman, Kahn 2008:1242, 1259
1324 Neumann 2009:35–38

Instrument und eine Methodik entwickelt, um Entwicklungsteams bei der Reflexion über normative Aspekte der Technologieentwicklung zu unterstützen. Die Autoren bezeichnen ihre Methode als ein Werkzeug zur Umsetzung von Ethics by Design, aber ich ordne ihre Methode unter Risikoanalysen ein, da seine Grundlage eine Folgenabschätzung ist. Folgenabschätzungen (Impact Assessments) für neue Technologien sind laut Urquhart und Craigon ein beliebtes Instrument zur Vorhersage von Problemen und zur strukturierten Entwicklung von Handlungsstrategien. Die Autoren schlagen zur Durchführung einer solchen Folgenabschätzung die Nutzung eines Spielbretts vor, das sie Moral-IT Impact Assessment Board nennen. Für die Folgenabschätzung wurden vier relevante Phasen identifiziert: Erstens die Ermittlung möglicher Risiken; zweitens die Bewertung der Bedeutung jedes der Risiken und der Wahrscheinlichkeit seines Eintretens; drittens die Festlegung geeigneter Schutzmaßnahmen für diese Risiken; und viertens die Untersuchung der praktischen Umsetzungsprobleme.[1325] Diese vier Phasen sollen anhand der zu entwickelnden Technologie durchgespielt werden, sodass geeignete Antworten und Aktionspläne ausgearbeitet werden können.[1326]

Dabei helfen soll das Moral-IT Kartendeck. Hierbei handelt es sich um einen Satz physischer Karten mit Fragen zu der Technologie, welche die Werte und Praktiken hinterfragen, die in technologische Systeme eingebettet sind. Der Text auf den Karten soll provozieren, um einen Ausgangspunkt für Diskussionen bilden zu können. Die Karten stellen ein Reflexionsinstrument dar, welches Entwicklungsteams dabei unterstützen soll, sich ihrer ethischen Verantwortung bewusst zu werden. Die Value-Sensitive Design Methode hat auf der Suche nach Umsetzungswegen die Gestaltung der Karten beeinflusst, und unterstützt damit Entwicklungsteams beim Nachdenken über Werte im Design.[1327] Das Moral-IT Kartendeck und das Moral-IT Spielbrett zur Folgenabschätzung unterstützen auch die Methode der Responsible Innovation. Denn in einem deliberativen Prozess (deliberation) werden ethische Fragen antizipiert (anticipation), zum Nachdenken über deren Auswirkungen angeregt (reflection) und zum Antworten und Pläneformulieren aufgefordert

1325 Urquhart, Craigon 2021:101
1326 Ebd.:95
1327 Ebd.:94–104

(responsiveness).[1328] Das Neue an ihrer Methode ist laut Urquhart und Craigon die Kombination der Folgenabschätzung mit einem Reflexionsinstrument wie dem Moral-IT Kartendeck. Hierdurch soll es möglich sein, während der Planung konkurrierende Interessen auszugleichen und schwierige moralische Entscheidungen zu treffen.[1329]

Das Moral-IT Kartendeck und das Spielbrett für die Folgenabschätzung bieten konkrete praktische Anleitung (i). Laut Urquhart und Craigon sind es die Entwicklungsteams, die sich ihrer ethischen Verantwortung bewusst werden und mit Hilfe der Folgenabschätzung über Werte in der Gestaltung von Technologien nachdenken sollten (iii). Aber es bleibt unklar, ob der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird (ii), denn die Folgenabschätzung kann kritisiert werden als Methode, die sich ausschließlich auf die Risikobewertung stützt und die potenziellen Auswirkungen auf die Gesellschaft oder die Umwelt kaum berücksichtigt.[1330] Laut Hermansson und Hansson liefert die klassische Risikoanalyse nicht alle notwendigen Informationen, denn sie fokussiert sich auf Wahrscheinlichkeiten und Schweregrade von Ergebnissen und missachtet ethische Aspekte des Handelns und der zwischenmenschlichen Beziehungen: „In order to appraise an action of risk-taking or risk imposition from a moral point of view, it is not sufficient to know the values and probabilities of its possible outcomes."[1331] Bei der Folgenabschätzung fehlt die Information, wer die Handlung durchführt und mit welchen Absichten.[1332] Denn wenn Risiko und Nutzen nur anhand der Wahrscheinlichkeit des Eintretens und der Schwere der Konsequenzen abgewogen werden, ist es möglich, dass einzelne Menschen bei der Vorkehrung von Risiken als finanzielle Bürde bewertet werden und nicht der Wert eines jeden betroffenen Menschen betrachtet wird.[1333] Da somit das allgemeine Kriterium ii) nicht erfüllt wird, ist es obsolet die spezifischen Kriterien zu betrachten und somit weist auch die Folgenabschätzung gemäß meiner Kriterien Mängel auf.

1328 Owen et al. 2013:44; Urquhart, Craigon 2021:97
1329 Urquhart, Craigon 2021:95, 97
1330 Owen et al. 2013:41
1331 Hermansson, Hansson 2007:130
1332 Hansson 2018:1820f.; Hermansson, Hansson 2007:130
1333 Lincke 2016:78

9.3.2 Internes Audit

Audits sind Instrumente zur Überprüfung komplexer Prozesse. Hierdurch kann festgestellt werden, ob Prozesse mit den Unternehmensrichtlinien, Industrienormen oder Vorschriften übereinstimmen. Die Informatikerin und Aktivistin Inioluwa Deborah Raji und ihre Kolleg:innen schlagen vor, interne Algorithmus-Audits zu nutzen, um zu überprüfen, ob die Prozesse der Entwicklung von KI-Systemen den erklärten ethischen Erwartungen und Standards der Organisation entsprechen. Mit Hilfe eines solchen Audits können potenzielle negative Folgen im Vorhinein antizipiert werden. Anschließend bieten die Auditergebnisse Entscheidungshilfen für die Entwicklung von Abhilfemaßnahmen oder auch für den Abbruch der Technologieentwicklung, falls die Risiken den Nutzen überwiegen. Interne Audits werden während der Produktentwicklung und vor der Markteinführung durchgeführt. Die Durchführung wird angeleitet durch ein Audit-Team, aber auch das Produktteam, die Geschäftsleitung und andere Stakeholder sind an dem Audit beteiligt.[1334]

Das interne Audit dient der Bewertung und Optimierung im Hinblick auf den gesellschaftlichen Nutzen und erklärte ethische Erwartungen und Standards der Organisation. Diese können zum Beispiel KI-Grundsätze sein, denen sich das betreffende Unternehmen verschrieben hat.[1335] Interne Audits überprüfen die Übereinstimmung einer Technologie mit den erklärten KI-Grundsätzen noch vor ihrem Einsatz. Das funktioniert, indem eine Risikoanalyse durchgeführt wird, welche fragt, was bei dem Einsatz der Technologie misslingen könnte.[1336] Raji et al. wenden jedoch ein: „Risk-based assessments can be limited in their ability to capture social and ethical stakes, and they should be complemented by anticipatory

1334 Raji et al. 2020:33f.; Raji et al. reflektieren, dass interne Auditor*innen zwangsläufig ein Interesse haben an einem guten Verlauf der Prüfung ihrer Organisation. Sie empfehlen, dass sich die Auditor*innen ihrer eigenen Vorurteile und Standpunkte sowie derjenigen der Organisation bewusst werden, um zu vermeiden, dass die Prüfungen lediglich dem Reputationsmanagement dienen. Außerdem sollen deshalb laut Raji et al. interne Prüfungen nur ein Aspekt eines umfassenderen Systems von Qualitätskontrollen sein (Raji et al. 2020:42).

1335 Raji et al. 2020:34f.

1336 Ebd.:35, 37

questions such as, 'what if…?'."[1337] Deshalb ergänzen die Autor:innen die formale Risikobewertung um die Kernanforderung der Antizipation aus der Responsible Innovation Methode. Das Ziel ist dabei, die ethische Vorausschau zu verbessern, indem systematisch nachgedacht wird über das sozio-technische System, in dem ein Produkt eingesetzt wird.[1338]

Der Rahmen für ein internes Audit, den Raji et al. entwerfen, besteht aus fünf Phasen: „(...) Scoping, Mapping, Artifact Collection, Testing and Reflection (SMACTR) (...)."[1339] In der Scoping-Phase werden Anwendungsfälle definiert sowie die Grundsätze und Werte bestätigt, die die Produktentwicklung leiten sollen. Zudem wird zum einen eine ethische Überprüfung der Anwendungsfälle durchgeführt, um die Frage zu beantworten, wer wahrscheinlich von der auditierten Technologie betroffen sein wird und auf welche Weise. Zum anderen werden die sozialen Auswirkungen bewertet (social impact assessment), indem die Schwere der Risiken eingeschätzt wird und die relevanten sozialen, wirtschaftlichen und kulturellen Auswirkungen und Schäden identifiziert werden.[1340] In der Mapping-Phase wird das Systemumfeld analysiert, werden Beteiligte dokumentiert und mithilfe der Fehlermöglichkeits- und Einflussanalyse (FMEA) potenzielle Fehlerquellen im Design identifiziert.[1341] Um Transparenz im Entwicklungsprozess zu schaffen und eine fundierte Prüfung zu ermöglichen wird in der Artifact Collection Phase dokumentiert, wie das zu auditierende Modell erstellt wurde, welche Annahmen während der Entwicklung getroffen wurden und welche Arten von Modellverhalten bei verschiedenen kulturellen, demografischen oder phänotypischen Gruppen auftreten könnten, wie die Daten erhoben wurden und ob sie sich auf Personen beziehen.[1342] Die Testing-Phase dient dazu, mit Hilfe von Simulationen und absichtlicher Fehlerauslösung zu prüfen, ob das System mit ethischen Werten vereinbar ist.[1343] In der Reflection-Phase wird abschließend eine Risikoanalyse in Bezug auf die algorithmische Nutzung (Algo-

1337 Raji et al. 2020:38
1338 Ebd.:38
1339 Ebd.:37
1340 Ebd.:39
1341 Ebd.:36
1342 Ebd.:40f.
1343 Ebd.:41

rithmic Use-related Risk Analysis) durchgeführt, welche sich auf die soziale Folgenabschätzung (social impact assessment) aus der Scoping Phase stützen sollte. Außerdem werden Maßnahmen zur Risikominderung entwickelt und alle Nachweise für die Entwicklung des Produktes in Übereinstimmung mit den ethischen Werten der Organisation für den Auditbericht dokumentiert.[1344]

Dadurch, dass die Risikoanalyse aus einer sozialen Folgenabschätzung (social impact assessment) besteht, wird deutlich, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird und somit der Vorteil für alle von der Technologie betroffenen Personen das Ziel sein soll (ii). Die Verantwortung für die Durchführung des internen Audits liegt eindeutig bei dem Audit-Team (iii). Doch die Verantwortung für das Vorlegen einer Erklärung oder Bestätigung der ethischen Ziele, Standards und KI-Prinzipien liegt beim Produkt- und Entwicklungsteam. Es wird aber nicht spezifiziert, wie diese Teams die ethischen Ziele, Standards und KI-Prinzipien identifizieren sollen. Ohne eine Anleitung lastet diese schwierige Aufgabe auf den Schultern von Mitarbeitenden, die auch noch andere Aufgaben haben und vermutlich keine Erfahrung haben im Identifizieren von ethischen Zielen. In ihrem Beispiel haben sich Raji et al. an den fünf am häufigsten genannten KI-Prinzipien aus dem Literatur Review von Jobin et al. orientiert.[1345] Meine Kritik an diesem Vorgehen ist der Einwand, dass die Häufigkeit der Nutzung bestimmter Prinzipien nicht allein ihre Begründung darstellen sollte.

Hier könnte man meinen, dass die ethischen Überlegungen dieser Arbeit wieder die Lösung sein könnten, wie bei den Ethischen Leitlinien und beim Value-Sensitive Design. Aber beim internen Audit geht es nicht darum konzeptionell alle für die Technologie und ihren Kontext wichtigen ethischen Werte zu identifizieren, sondern es geht darum die ethischen Werte der Organisation festzustellen und die Technologie daran auszurichten. Das heißt es bedarf eines Reflektionsprozesses innerhalb der betreffenden Organisation über die eigenen ethischen Ziele, Standards und KI-Prinzipien. Das interne Audit von Raji et al. bietet durch die klar beschriebenen fünf Phasen (SMACTR) konkrete praktische Anleitung (i), aber es fehlen

1344 Raji et al. 2020:41f.
1345 Jobin et al. 2019; Raji et al. 2020:38

Instruktionen für das Produkt- und Entwicklungsteam, wie es zu einer Erklärung oder Bestätigung der ethischen Ziele, Standards und KI-Prinzipien der betreffenden Organisation kommen kann.[1346]

Wenn man die Methode des internen Audits uminterpretiert und dafür nutzt zu prüfen, ob die Technologie mit allgemeingültigen ethischen Werten übereinstimmt, anstatt mit den ethischen Werten der Organisation, könnten die Werte mit Hilfe der in dieser Arbeit entwickelten deontologischen Risikoethik begründetet werden und es wären damit alle allgemeinen Kriterien erfüllt. Trotzdem würde die Methode des internen Audits an mindestens einem der spezifischen Kriterien für die ethische Gestaltung eines Datenökosystems scheitern: es wird nicht ermöglicht, die Auferlegung von Risiken zu regulieren (I). Das interne Audit betrachtet in der Scoping-Phase ausführlich wer wahrscheinlich von der auditierten Technologie betroffen sein wird und auf welche Weise, wie schwer die Risiken sind und welche sozialen, wirtschaftlichen und kulturellen Auswirkungen sie haben können und führt in der Reflection-Phase auf dieser Grundlage eine Risikoanalyse durch – dadurch wird die Risikoauferlegung problematisiert. Aber dabei wird nicht klar, unter welchen Umständen das Risiko zulässig ist[1347], sodass die Auferlegung von Risiken nicht reguliert werden kann. Da somit Kriterium I) nicht vollständig erfüllt werden kann, ist auch das interne Audit nicht ausreichend zur Umsetzung von ethischen Überlegungen in die Praxis.

9.3.3 Ethische Risikoanalyse

Hansson hat, angelehnt an das Drei-Parteien-Modell aus Kapitel 3.2.1, das er mit Hermansson entwickelt hat, eine ethische Risikoanalyse (eRA) entwickelt. Er kritisiert, dass der Fokus bei der modernen Risikoanalyse fast ausschließlich auf den Wahrscheinlichkeiten und der Schwere der unerwünschten Folgen liegt. Seine

1346 Raji et al. 2020:39

1347 Raji et al. schlagen vor eine Risikoschwelle zu definieren, unterhalb derer das Risiko als tolerierbar angesehen wird (Raji et al. 2020:42), aber es bleibt unklar, wie diese Schwelle definiert werden könnte. Außerdem wäre es, wie unter 4.3 erläutert, laut Hansson sehr schwierig, diese Grenze zu ziehen, weil eine pauschale Risikogrenze nicht in der Lage wäre, die Vorteile der jeweiligen risikobehafteten Handlung einzubeziehen (Hansson 2003:299).

eRA hingegen soll stattdessen die Fragen des Handelns und der zwischenmenschlichen Beziehungen betrachten, die den Kern von ethischen Diskussionen über Risiken bilden. Dabei soll die eRA von unparteiischen Akteuren durchgeführt werden, wie zum Beispiel akademischen, gemeinnützigen und in einigen Ländern staatlichen Organisationen. Um Risiken auf ethische Art und Weise zu analysieren, werden drei Schritte durchlaufen: Zuerst werden die Personen identifiziert, die eine primäre ethisch relevante Rolle in Bezug auf das Risiko haben. Dann werden ihre Rollen, Rollenkombinationen, Wechselbeziehungen und potenziellen Konflikte beschrieben. Abschließend werden diese Konflikte mit einem Schwerpunkt auf Fragen nach Gerechtigkeit, Rechten und sozialer Macht analysiert.[1348]

Hansson kritisiert die Ansätze und Methoden, welche das Einbeziehen von Stakeholdern empfehlen (z.B. Human-centered Design, Responsible Innovation, Value Sensitve Design), da seiner Meinung nach Stakeholder aus instrumentellen oder strategischen Gründen in die Analyse einbezogen werden und nicht aus ethischen Gründen. Zudem unterscheidet die Stakeholder-Analyse seiner Ansicht nach nicht zwischen verschiedenen Gruppen von Stakeholdern: „From an ethical point of view, the large differences among stakeholders in terms of risk exposure, power, and resources have to be highlighted."[1349] Außerdem nimmt die Stakeholder-Analyse typischerweise eine Insider-Perspektive ein, während eine ethische Risikoanalyse laut Hansson eine unparteiische Outsider-Perspektive einnehmen sollte. Und die Stakeholder-Analyse versucht einen Konsens zwischen den Beteiligten zu erzielen, während eine ethische Analyse auf die unterschiedliche Verhandlungsmacht der verschiedenen Stakeholder aufmerksam macht und nicht unbedingt einen Konsens anstrebt.[1350]

In dem ersten Schritt einer ethischen Risikoanalyse müssen die verschiedenen Stakeholdergruppen identifiziert werden. Dabei gibt es laut Hansson drei unterschiedliche Formen der unmittelbaren Beteiligung von Menschen an einer Risikosituation, die schon aus dem Drei-Parteien-Modell aus Kapitel 3 bekannt sind: die Risikoexponierten, die Entscheidungstragenden und die durch das Risiko Be-

1348 Hansson 2018:1820f.
1349 Ebd.:1821
1350 Ebd.:1821f.

günstigten. Risikoexponierte können nicht immer individuell identifiziert werden, aber Gruppen, die von dem Risiko betroffen sind, können in die eRA miteinbezogen werden. Dabei sind die Gruppen in unterschiedlicher Weise und Höhe dem Risiko ausgesetzt: „They may also differ widely in their awareness of the risk, their capacity to make themselves heard, and the resources they can muster for furthering their interests."[1351] Dadurch, dass die Gruppen der Risikoexponierten identifiziert werden, kann sichergestellt werden, dass ihre Rechte und Interessen beachtet werden.

Eine Situation, die für die einen ein Risiko birgt, kann für andere von Vorteil sein. Es ist wichtig, diese Gruppe von Begünstigten zu identifizieren, um festzustellen, warum das Risiko in Kauf genommen wird. Zusätzlich werden im ersten Schritt der eRA die Entscheidungstragenden identifiziert: „(…) we are looking for those persons who (knowingly or due to ignorance for which they can be held responsible) make decisions that contribute causally to a risk exposure."[1352] Dies kann schwierig sein, da auch Hintergrundentscheidungen, zum Beispiel durch externe Hersteller und Regulierungsbehörden, die Bedingungen schaffen, welchen die Risikoexponierten ausgesetzt sind. Durch solche Hintergrundentscheidungen haben Risikoexponierte dann eventuell nur begrenzte Möglichkeiten, sich gegen das Risiko zu entscheiden.

Im zweiten Schritt der eRA wird betrachtet, ob manche Akteure mehrere dieser Rollen einnehmen, ob sich verschiedene Akteure gegenseitig beschränken und ob sie voneinander abhängig sind. Dabei stellt sich heraus, dass bestimmte Rollenkombinationen problematischer sind als andere. So ist es zum Beispiel ethisch hochproblematisch, wenn ein Akteur ausschließlich die Rolle der Risikoexponierten erfüllt, während andere Akteure über das Risiko entscheiden und davon profitieren. Hansson begründet das so: „(…) a disadvantage to one person is not necessarily outweighed by a larger advantage to someone else."[1353] Selbst wenn bestimmte Akteure einen großen Vorteil von dem Risiko haben, ist es nicht unbedingt ethisch gerechtfertigt, dieses Risiko anderen Akteuren aufzuerlegen.

Weniger problematisch wird es, wenn die Risikoexponierten gleichzeitig auch die durch das Risiko Begünstigten sind. Hier sind

1351 Hansson 2018:1822
1352 Ebd.:1823
1353 Ebd.:1826

die Ergebnisse vorteilhaft, solange der Nutzen einen möglichen Schaden überwiegt. Ein Nachteil ist jedoch, dass die Entscheidung, welche diese Risikoexponierten betrifft, immer noch von anderen getroffen wird. Ein Beispiel dafür wäre eine paternalistische Entscheidung im medizinischen Kontext: Bewusstlose Patient:innen profitieren von einer medizinischen Maßnahme und tragen das Risiko, dass die erwünschte Wirkung ausbleibt, aber die Entscheidung treffen Ärzt:innen.

Eine andere Möglichkeit wäre es, dass die Risikoexponierten auch gleichzeitig die Entscheidungstragenden sind und zum Beispiel in einen Fluss springen, um eine andere Person vor dem Ertrinken zu retten. So lange sich der Akteur frei entscheiden kann, das Risiko einzugehen, ist diese Rollenkombination ethisch nicht problematisch – auch wenn der Akteur nicht selbst von dem Risiko profitiert.

Es ist sogar möglich, alle drei Rollen zu verkörpern. Zum Beispiel, wenn ein Akteur entscheidet, abseits der Pisten Ski zu fahren, nimmt er das erhöhte Risiko eines Sturzes aktiv auf sich, ist aber auch derjenige, der davon profitiert. Die Kombination aller drei Rollen klingt vorerst ideal, aber es können trotzdem noch andere Akteure in negativer Weise betroffen sein. Beim Skifahren abseits der Piste kann zum Beispiel eine Lawine ausgelöst werden, die andere Menschen einem Risiko aussetzt, das sie nicht gewählt haben und von dem sie nicht profitieren.

Außerdem können andere Entscheidungstragende die Möglichkeiten des Akteurs beschränken. So stellt sich zum Beispiel die Frage, welche Verantwortung eine Person trägt, die nicht alle Pisten so abgesperrt hat, dass niemand abseits der Piste Skifahren kann. Diese Person wäre in dem Fall ausschließlich in der Rolle eines Entscheidungstragenden. Ihre Aufgabe wäre es, unparteiisch zu sein und die Interessen der Risikoexponierten und Begünstigten auszubalancieren. Dies ist eine verantwortungsvolle Aufgabe, da die von der Entscheidung betroffenen Personen die Entscheidung nicht beeinflussen können.

Auch wenn Akteure nur die Rolle von Begünstigten verkörpern, muss über die Verteilung der Vor- und Nachteile nachgedacht werden. Für den begünstigten Akteur ist dies eine vorteilhafte Position, aus gesellschaftlicher Sicht erscheint das jedoch unfair für diejenigen, die das Risiko tragen. Wenn aber ein Akteur sowohl zu den Entscheidungstragenden als auch zu den Begünstigten gehört, ist das

die problematischste Rollenkombination aus ethischer Sicht. Dieser Akteur entscheidet über das Risiko und profitiert davon, dass es eingegangen wird, aber geht das Risiko nicht selbst ein. Dadurch entsteht der Anreiz, anderen mehr Risiken aufzuerlegen, um selber davon zu profitieren, ohne Nachteile dadurch zu erleiden.[1354]

Im dritten Schritt der eRA können verschiedene Instrumente zur ethischen Analyse der jeweiligen Rollenkombinationen eingesetzt werden. Hansson schlägt vor, dass für Akteure, die sowohl dem Risiko ausgesetzt sind, als auch von ihm profitieren, eine individuelle Abwägung zwischen Risiko und Nutzen durchgeführt wird. Eine solche Abwägung kennt man zum Beispiel aus dem medizinischen Kontext, in welchem im Idealfall die Patientin selbst die Risiken und den Nutzen einer bestimmten Behandlung abwägt, um möglichst ein Gleichgewicht zu erreichen, das auf den eigenen Werten und Prioritäten basiert.[1355]

Für den Fall, dass diese individuelle Analyse eine ungleiche Verteilung zwischen Risiko und Nutzen ergibt, schlägt Hansson die Durchführung einer Verteilungsanalyse vor. Für eine Verteilungsanalyse müssen quantitative Schätzungen oder Indikatoren sowohl für die Risiken als auch für den Nutzen erstellt werden. Mit diesen soll festgestellt werden, wie die Verteilung der neuen Risiken und Nutzen mit bereits bestehenden Ungleichheiten zusammenhängt. So zeigt sich, ob die Rollenverteilung fair ist. Denn es zeigt sich, ob die Risiken vorrangig von sowieso schon benachteiligten Gesellschaftsgruppen getragen werden und ob sowieso schon privilegierte Gesellschaftsgruppen davon profitieren. Laut Hansson ist die Durchführung dieser Verteilungsanalyse auch bei allen anderen Rollenkonstellationen sinnvoll, außer wenn ein Akteur ausschließlich die Rolle der Entscheidungstragenden einnimmt.[1356]

Wenn nun das Risiko nicht durch den Nutzen aufgewogen wird oder wenn eine Gruppe von Risikoexponierten niemals von dem Risiko profitiert, sollte laut Hansson analysiert werden, ob ihre moralischen Rechte verletzt wurden. Denn, wie in Kapitel 5 ausführlich besprochen, ist es laut Hansson nur dann zulässig, ein Risiko von außen aufzuerlegen, wenn dies Teil eines fairen und für beide

1354 Hansson 2018:1824ff.
1355 Ebd.:1826
1356 Ebd.:1827

Seiten vorteilhaften Austauschs von Risikoexpositionen ist.[1357] Insbesondere, wenn es machtlose, risikoexponierte Personen gibt oder Entscheidungstragende, die davon profitieren, dass andere einem Risiko ausgesetzt sind, dann ist laut Hansson zusätzlich eine Analyse der zugrunde liegenden Machtstruktur erforderlich. Er schlägt vor, diese Frage zu stellen: Was für ein System der Entscheidungsfindung für diese Art von Risiko würden wir aufbauen, wenn wir von vorne anfangen würden? Auf dieser Grundlage können Vor- und Nachteile des aktuellen Systems diskutiert und Alternativen vorgeschlagen werden.[1358]

Schritt drei der eRA kann außerdem die Identifikation von weiteren für die Risikoanalyse relevanten Akteuren beinhalten. So können zum Beispiel wissenschaftliche Expert:innen oder Journalist:innen als unabhängige Informationslieferant:innen fungieren, und Personen wie Rechtsanwält:innen und Lobbyist:innen können als Vertreter:innen von Risikoexponierten, Entscheidungstragenden oder Begünstigten auftreten. Wissenschaft spielt eine Rolle, wenn es darum geht, Risiken zu erkennen, zu bestätigen und ihr Ausmaß zu bestimmen. Journalist:innen spielen eine Rolle, da die Öffentlichkeit häufig vor allem über die Medien über potenzielle Risiken informiert wird. Problematisch wird es, wenn Wissenschaftler:innen oder Journalist:innen Verbindungen zu mächtigen, durch das Risiko begünstigten Akteuren haben. Deshalb bewertet die eRA, ob jegliche Informationslieferant:innen den von ihnen erwarteten ethischen Grundsätzen gerecht werden. Bei dem Einsatz von Vertreter:innen ist es die Aufgabe der eRA zu überprüfen, ob auch hier moralisch zulässig gehandelt wird.

Die ethische Risikoanalyse bietet dank ihrer drei Schritte eine konkrete praktischen Anleitung für die Ausrichtung technischer Entwicklungen an ethischen Maßstäben (i). Durch den Fokus auf die Fragen, wer durch das Risiko begünstigt wird, wird deutlich, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird und der Vorteil für alle von der Technologie betroffenen Personen das Ziel ist (ii). Die Verantwortung für die ethische Gestaltung wird unparteiischen Akteuren zugeschrieben, damit eine Outsid-

1357 Hansson 2018:1827
1358 Ebd.

er-Perspektive eingenommen werden und auf die unterschiedliche Verhandlungsmacht der verschiedenen Stakeholder aufmerksam gemacht werden kann (iii). Bei den unparteiischen Akteuren kann es sich zum Beispiel um akademische, gemeinnützige oder staatliche Organisationen handeln. Verantwortlich sollen also menschliche Akteure sein, aber es bleibt vage, welche bestimmten menschlichen Akteure die Verantwortung übernehmen sollen, aber es werden zumindest mehrere Vorschläge gemacht, die sich noch ausgestalten ließen. Somit erfüllt die ethische Risikoanalyse alle allgemeinen Kriterien für die ethische Gestaltung einer Technologie.

Konkret in Bezug auf die ethische Gestaltung eines Datenökosystems wäre es mit Hilfe der eRA zumindest möglich, Verletzung von Privatsphäre und Abwesenheit von Transparenz als Risiken zu identifizieren. Denn dank der eRA könnte man erkennen, dass Datenquellen das Risiko der Verletzung ihrer Privatsphäre und Datennutzenden das Risiko fehlender Transparenz auferlegt wird. Da Datenquellen und Datennutzende nicht über das ihnen auferlegte Risiko entscheiden können, handelt es sich bei diesem Akteur um machtlose, risikoexponierte Personen. Deshalb wäre hier laut Schritt drei der eRA eine Analyse der zugrunde liegenden Machtstruktur erforderlich.[1359] Sie schlägt aber keine konkreten Schritte vor, welche die Ableitung von administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen (II) und zur Umsetzung von Transparenz für die Datennutzenden (III) ermöglichen würde, sodass auch ihre Erfüllung schwierig wäre.

Ich stimme Hansson zu, dass beim Vorkommen von „powerless risk-exposed persons“[1360] die Machtverhältnisse, bzw. wie ich sagen würde die Möglichkeiten zur Kontrolle, analysiert werden sollten. Ich denke aber, dass es nicht reicht, sich das ideale System auszumalen und Alternativen zum aktuellen System zu diskutieren. Das hieße im Kontext von Verwaltungsleistungen, dass man das gesamte deutsche Verwaltungssystem hinterfragen müsste – das würde nur unser System destabilisieren und nicht das Problem der fehlenden Risiko-Kontrolle lösen. Ich argumentiere, dass wir stattdessen Bedingungen definieren müssen, unter welchen es zulässig ist, anderen ein Risiko aufzuerlegen.

1359 Hansson 2018:1827
1360 Ebd.

Meiner Meinung nach müsste also in Schritt drei der eRA analysiert werden, unter welchen Umständen das Risiko zulässig ist. Denn nur so ist es möglich, die Auferlegung von Risiken zu regulieren (I). Laut Hansson ist es nur dann zulässig, ein Risiko von außen aufzuerlegen, wenn dies Teil eines fairen und für beide Seiten vorteilhaften Austauschs von Risikoexpositionen ist.[1361] Ich habe jedoch in Kapitel 5 argumentiert, dass es für eine zulässige Risikoauferlegung nicht reicht, wenn ein Risiko durch einen eigenen Vorteil kompensiert wird. Meine Ableitung einer Regel für fairen und für beide Seiten vorteilhaften Austausch von Risikoexpositionen aus Kants dritter Formel des Kategorischen Imperativs kann an dieser Stelle klarere Antworten geben und die Regulierung von Risikoauferlegung ermöglichen. Deshalb entwickle ich im Folgenden auf Basis der eRA sowie der ethischen Überlegungen in dieser Arbeit eine eigene Umsetzungsmethode zur ethischen Gestaltung eines Datenökosystems.

1361 Hansson 2018:1827.

10 Eigener ethischer Verfahrensmaßstab für die Gestaltung von Datenökosystemen

Da keiner der beleuchteten Ansätze alle Kriterien für die ethische Gestaltung von Technologien im Allgemeinen und von Datenökosystemen im Besonderen erfüllt, entwickle ich in diesem Kapitel einen eigenen ethischen Verfahrensmaßstab. Dieser soll durch eine konkrete praktische Anleitung ermöglichen, technische Entwicklungen an ethischen Maßstäben auszurichten. Die ethischen Anforderungen an die Gestaltung von Datenökosystemen werden dabei aus der in dieser Arbeit entwickelten deontologischen Risikoethik abgeleitet. Bei jedem vorgeschlagenen Schritt soll der Vorteil für alle von der Technologie betroffenen Personen das Ziel sein. Außerdem wird die Verantwortung für die ethische Gestaltung des Datenökosystems einem bestimmten kollektiven Akteur zugeordnet.

Ich schlage vor, dass es ein Ethikteam geben sollte innerhalb der Unternehmen, die Datenökosysteme entwickeln. Es ist jedem Unternehmen selbst überlassen, ob es seine Produkte ethisch gestalten möchte oder nicht. Denn die Ethik hat keinen Durchsetzungsmechanismus, wie es das Recht hat. Es sollte aber im Interesse des gesamten Unternehmens sein, seine Produkte ethisch zu gestalten, um so möglichen negativen Konsequenzen vorzubeugen und eine aktive gute Gestaltung zu ermöglichen – dafür trägt das Unternehmen als kollektiver Akteur die moralische Verantwortung. Jedes Unternehmen sollte sich dieser Verantwortung bewusst werden und dementsprechend handeln, indem ein Ethikteam eingerichtet und ernst genommen wird.

Mitglied dieses Teams sollte mindestens ein:e Ethiker:in sein, welche:r erklären kann, woher die ethischen Anforderungen stammen und warum es notwendig ist, den ethischen Verfahrensmaßstab anzulegen und die Maßnahmen, die sich daraus ergeben, durchzuführen. Das Ethikteam sollte außerdem aus Mitarbeitenden bestehen, die das zu gestaltende Produkt kennen und seine Auswirkungen

abschätzen können. Beteiligt sind hier also Vertreter:innen der Geschäftsführung eines Unternehmens (Strategisches Management), welche beschließt, das Produkt herzustellen, des Projektteams (Konzeptentwicklung), welches über die genauen Spezifikationen des Produkts entscheidet (Konzeptentwicklung), und der Entwickler:innen (Entwicklungsprozess), welche das Produkt innerhalb der Grenzen der Spezifikationen gestalten.[1362]

Die Verantwortung für die ethische Gestaltung des gesamten Datenökosystems bzw. jeder der eingesetzten Technologien fällt dem gesamten Unternehmen als kollektivem Akteur zu, da jede der drei Ebenen Kontrolle ausübt über die Gestaltung oder zumindest eine Beaufsichtigungsrolle einnimmt. Aber die Verantwortung für die Einhaltung des ethischen Verfahrensmaßstabs wird dem Ethikteam übertragen. Es sind also ausgewählte Mitarbeitende, geleitet durch Ethiker:innen dafür verantwortlich, sich konkret mit den ethischen Bedenken auseinanderzusetzen und Lösungsvorschläge zu erarbeiten. Im besten Fall entscheiden sich die Mitarbeitenden aus eigenem Interesse für ihre Teilnahme am Ethikteam, weil sie besonderen Wert darauflegen, dass die ethischen Bedenken bezüglich ihrer Produkte durchdacht und minimiert werden. Sie sollten für regelmäßige Treffen des Ethikteams freigestellt werden und sowohl Ergebnisse als auch offene Probleme zurück in ihren Bereich tragen und umgekehrt – so können auch Mitarbeitende außerhalb des Ethikteams ihre Lösungsvorschläge einbringen.

Eine solche ethische Gestaltung ist möglich, wenn das Ethikteam während seiner Treffen den im Folgenden beschriebenen vier Schritten folgt. Die vier Schritte sollten schon im Planungsprozess eines neuen Datenökosystems durchlaufen werden. Sie können aber auch genutzt werden, um ein Datenökosystem auf ethische Probleme zu überprüfen, während es schon im Einsatz ist. Die Ethiker:innen sollten das Vorgehen des Ethikteams dokumentieren und am Ende darüber einen Abschlussbericht schreiben, damit zumindest intern eine Überprüfung möglich ist. Die durchzuführenden Schritte umfassen zuerst die Identifikation von Nutzen und Risiken der zu gestaltenden Technologie, dann folgt Hanssons ethische Risikoanalyse (eRA). Diese wird noch ergänzt um die Antwort auf die Frage, unter welchen Bedingungen, außer einem gegenseitigen Vorteil, Ri-

1362 Gogoll et al. 2021:1087

siken zulässig sind, indem Kants dritte Formel des Kategorischen Imperativs als eine Regel eines sozialen gerechten Systems der Risikobereitschaft aufgestellt wird. Zum Schluss werden Maßnahmen erforderlich, welche die identifizierten Risiken abmildern – insbesondere, wenn Akteure, die weniger Kontrolle über sie haben, von ihnen betroffen sind.

Schritt 1: Positionierung zu Verantwortung, Nutzen und Risiken

Zuerst sollte sich das Ethikteam bewusst machen, wer verantwortlich ist für die ethische Gestaltung der betreffenden Technologien bzw. des gesamten Datenökosystems. Das gelingt, indem identifiziert wird, wer in dem zu bewertenden Fall die Datengenerierenden und Datenverarbeitenden sind. Es ist wahrscheinlich, dass das Unternehmen, für welches das Ethikteam diesen Verfahrensmaßstab durchführt, mindestens eine dieser Rollen einnimmt – es ist möglich, dass es sogar beide Rollen verkörpert. Damit gehört natürlich auch das Ethikteam, als Teil des Unternehmens, zu diesem verantwortlichen kollektiven Akteur. Hier erwächst ein Bewusstsein für die epistemisch vorteilhefte Position, in der sich das Ethikteam befindet, sowie für ihre Kontrollmöglichkeiten über die Vorgänge im zu gestaltenden Datenökosystem. Als nächstes sollte sich das Ethikteam klar machen, dass es im Datenökosystem immer auch Datenquellen und Datennutzenden geben muss. Denn das macht darauf aufmerksam, dass man zur Gestaltung des Datenökosystems, und seiner einzelnen Technologien, Daten braucht. Im Folgenden wird angenommen, dass es sich um ein Datenökosystem handelt, das personenbezogene Daten erfordert, sodass diese von Personen (Datenquellen) zur Verfügung gestellt werden müssen. Außerdem wird hierdurch klar, dass am Ende jemand die Ergebnisse nutzen können soll (Datennutzende). Daraufhin sollen Nutzen und Risiken der zu gestaltenden Technologie identifiziert werden.

Hier werden Fragen gestellt, die schon vor dem Beschluss, das betreffende Produkt herzustellen, beantwortet werden müssen: Welchem Zweck soll das Produkt dienen und was soll es zum Wohl der Beteiligten beitragen? Schon beim strategischen Management des Projektes muss klar werden, wofür einzelne Technologien oder ganze Datenökosysteme entwickelt werden. Die Befolgung eines Ge-

setzes, wie zum Beispiel dem OZG, kann ein Grund sein, sollte aber nicht der Einzige bleiben. Die Politiker:innen haben beim Erlassen des Gesetzes sicher erwartet, dadurch den Zugang zu Verwaltungsleistungen zu erleichtern. Aber wenn man selbst etwas zu diesem Zweck beiträgt, sollte man sich seine Gründe bewusst machen. Nur so ist man in der Lage festzustellen, ob der Nutzen die Risiken überwiegt.

Für eine solche Abwägung müssen die Risiken der Entwicklung und des Einsatzes des Produktes identifiziert werden. Hier sollte sich das Ethikteam zunächst Worst-Case-Szenarien vorstellen: Welche negativen Folgen könnte der Einsatz der zu gestaltenden Technologie haben? Eine Hilfestellung zur Identifikation von Risiken sollen die von mir aus der strukturellen Beschaffenheit von Datenökosystemen abgeleiteten Risiken sein: die Verletzung der Privatsphäre und das Fehlen von Transparenz. Sobald es sich um ein Produkt handelt, das mit dem Sammeln, Speichern und Weitergeben von (personenbezogenen) Daten zusammenhängt, spielt das Risiko für die Privatsphäre der Datenquellen eine Rolle. In gleicher Weise spielt das Risiko fehlender Transparenz für Datennutzende eine Rolle, sobald Daten mit Hilfe von Algorithmen verarbeitet und die Ergebnisse zur Nutzung bereitgestellt werden. Es sollten also die Privatsphäre der Datenquellen und die Transparenz für die Datennutzenden in Schritt 1 durch das Ethikteam herausgearbeitet werden. Darüber hinaus können auch zusätzliche Risiken identifiziert werden, die spezifisch für das betreffende Produkt sind.

Um schon im Vorhinein Verzerrungen durch die zu entwickelnde Technologie zu vermeiden, sollte sich das Ethikteam bei der Risikoidentifikation zudem an der Definition der drei Verzerrungen von Friedman und Nissenbaum orientieren. Es sollten erstens Vorurteile in Bezug auf kulturelle Identität, Klasse, Geschlecht, Lese- und Schreibfähigkeit sowie Behinderung reflektiert werden, um das Reproduzieren bestehender Verzerrungen (preexisting bias) zu vermeiden. Zweitens sollte die Nutzung des Systems und damit einhergehende technische Verzerrungen (technical bias) antizipiert werden, welche zum Beispiel durch Bildschirmgrößen oder durch systematische Ungleichbehandlung mittels eines Algorithmus entstehen können. Drittens sollte bedacht werden, in welchen Kontexten

das System genutzt werden könnte, um emergente Verzerrungen (emergent bias) zu minimieren.[1363]

Wenn sich schon nach Schritt 1 abzeichnet, dass das Risiko den Nutzen überwiegt, sollte in Frage gestellt werden, ob das betreffende Produkt überhaupt gestaltet werden sollte und das Ethikteam sollte sich an die Geschäftsführung wenden. Überwiegt aber der Nutzen die Risiken oder scheinen sich beide auszubalancieren, geht es weiter mit Schritt 2. Denn nach der Identifikation von Nutzen und Risiken kann nun betrachtet werden, wer davon betroffen ist.

Schritt 2: Risikoverteilung bewerten mit Hanssons ethischer Risikoanalyse

In Schritt 2 meines ethischen Verfahrensmaßstabs wird die ethische Risikoanalyse von Hansson durchgeführt. Das heißt, als erstes wird identifiziert, wer über das Risiko entscheidet, wer das Risiko trägt und wer davon profitiert.[1364] Diejenigen, die über das Risiko entscheiden, sind Teil des Ethikteams und sollten so Einfluss nehmen auf die Ausgestaltung der Risiken. Darüber hinaus gilt aber der kollektive Akteur – das gesamte Unternehmen – welcher das Datenökosystem entwickelt, als Entscheidungstragender und damit auch als rechtlich Verantwortlicher. Besonders wichtig ist also, dass das Ethikteam versteht, wer die Risikoexponierten und wer die durch das Risiko Begünstigten sind. Bei der Identifikation der Risikoexponierten kann sich das Ethikteam daran orientieren, wer die Datenquellen und Datennutzenden sind.

Als nächstes wird betrachtet, ob manche Akteure mehrere dieser Rollen einnehmen, ob sich verschiedene Akteure gegenseitig beschränken und ob sie voneinander abhängig sind.[1365] So kann zum Beispiel das Unternehmen als Entscheidungstragender gleichzeitig auch durch ein Risiko begünstigt sein. Wenn das Unternehmen dabei aber nicht zu den Risikoexponierten gehört, ist das ethisch problematisch, weil dies Anreize zur ungerechten Risikoabwälzung entstehen lässt. Sobald Nutzen und Risiken zwischen den Akteuren

1363 Friedman, Nissenbaum 1996:343f.; Friedman, Kahn 2008:1253
1364 Hansson 2018:1822
1365 Ebd.:1826

ungleich verteilt sind, empfiehlt Hansson eine Verteilungsanalyse durch quantitative Schätzungen oder Indikatoren, um die Rollenverteilung bewerten zu können.[1366]

Hanssons ethische Risikoanalyse hilft dabei zu verstehen, wie Risiken und Nutzen auf die verschiedenen Stakeholder aufgeteilt sind. Aber um zu verstehen, ob die Verteilung, und damit die Situation, welche durch die Gestaltung des Produktes entsteht, ethisch zulässig ist, muss es aus meiner Sicht einen weiteren Schritt geben: Es muss klar sein, unter welchen Bedingungen ein soziales System der Risikobereitschaft gerecht ist.

Schritt 3: Überprüfung der Zulässigkeit von Risikoauferlegung

Anhand dieser deontologischen Risikoethik kann das Ethikteam im dritten Schritt meines ethischen Verfahrensmaßstabs überprüfen, ob es ethisch zulässig ist, bei der Gestaltung eines Datenökosystems Akteuren ein Risiko aufzuerlegen. Demnach handelt es sich dann um ein gerechtes soziales System der Risikobereitschaft, wenn niemand als bloßes Mittel gebraucht wird. Da man freiwillig und informiert zustimmen können muss, um nicht als bloßes Mittel gebraucht zu werden, ist hier zu überprüfen, ob die Risikoexponierten über das Risiko informiert werden und ob sie die Möglichkeit haben zuzustimmen sowie abzulehnen. Dabei sollte sich das Ethikteam auch an meinem Prinzip der Maximierung der individuellen Kontrolle orientieren. Dies soll gewährleisten, dass die von Risiken Betroffenen, trotz des ihnen auferlegten Risikos, über Handlungsspielräume verfügen und nachvollziehen können, worauf sie sich einlassen. Eine solche Nachvollziehbarkeit setzt voraus, dass relevante Informationen in verständlicher Form bereitgestellt werden – ohne Zwang und ohne Irreführung. Die Bereitschaft, ein Risiko zu akzeptieren, ist dabei maßgeblich davon abhängig, ob die Betroffenen dadurch auch begünstigt werden und ob bei der Risikoauferlegung der Zweck der Achtung der Menschlichkeit gefördert wird.

Nur wenn niemand bloß als Mittel gebraucht wird, also jede:r die Möglichkeit der Kontrolle hat, findet das Auferlegen des Risikos im Rahmen eines sozialen gerechten Systems der Risikobereitschaft

1366 Hansson 2018:1827

statt und ist somit ethisch zulässig. Wenn es nicht möglich ist, das Risiko zur Wahl zu stellen, sollte das Risiko so gering wie möglich sein, und Risikoexponierte sollten für das Risiko sowie für mögliche Schäden Kompensation erhalten.[1367]

Schritt 4: Fehlervorsorge durch Minimierung der Risiken

Damit die Kontrolle der Risikoexponierten maximiert werden kann und somit die Risikoauferlegung zulässig wird, sollten im letzten Schritt des ethischen Verfahrensmaßstabs die Risiken so gestaltet werden, dass die Risikoexponierten nicht als bloßes Mittel gebraucht werden. Für die ethisch zulässige Gestaltung der Risiken einer verletzten Privatsphäre und fehlender Transparenz sollte sich das Ethikteam auf die administrativen Anforderungen beziehen, die in Kapitel 6 und 7 abgeleitet wurden. Denn nur wenn diese administrativen Anforderungen erfüllt sind, können Datenquellen und Datennutzende selbstbestimmt handeln, da ihre Privatsphäre und ihr Anspruch auf Transparenz sowohl geschützt als auch aktiv gefördert werden. In Bezug auf die Privatsphäre der Datenquellen geht es darum, beim Sammeln der Daten das Einverständnis der Datenquellen einzuholen, sie dabei vor unfreiwilliger Einwilligung zu schützen, indem man sie über jeden Verwendungszweck der Daten informiert, ihnen die Möglichkeit gibt zuzustimmen oder abzulehnen sowie zusätzlich die Daten vor Zweckentfremdung schützt. In Bezug auf die Transparenz für die Datennutzenden ist es wichtig, algorithmische Ergebnisse im zum Verständnis notwendigen Kontext darzustellen, sodass klar wird, wie die Ergebnisse entstanden sind und welche Aussagen sie enthalten.

Durch die Umsetzung der administrativen Anforderungen werden die Risiken minimiert. Hierfür sollte das Ethikteam kommunikative sowie technische Maßnahmen erarbeiten – für Letztere sind vor allem die Entwickler:innen des Ethikteams gefragt. Um zum Beispiel die Daten der Datenquellen vor Zweckentfremdung zu schützen, können Maßnahmen wie Datenminimierung oder Anonymisierung helfen. Wenn es darum geht, das freiwillige Einverständnis der Datenquellen einzuholen, indem man sie über jeden Verwendungs-

1367 Hansson 2003:303

zweck der Daten informiert, dann muss das auch technisch umsetzbar sein. Aber vor allem ist dabei wichtig, sich über die Kommunikation klar zu werden, die man mit den Datenquellen eingehen möchte. So ist auch für das Verständnis der algorithmischen Ergebnisse womöglich eine technische Lösung notwendig, wie zum Beispiel die Vermeidung von Black Box Algorithmen oder die Verwendung und Interpretation von Explainable AI. Wenn es aber darum geht, algorithmische Ergebnisse im zum Verständnis notwendigen Kontext darzustellen, ist vor allem wichtig, wie verständlich kommuniziert werden kann, welche Aussage die Ergebnisse enthalten.

Wenn die Risiken nicht abgeschwächt werden können oder es nicht durch technische Lösungen vermeidbar ist, dass Personen als bloßes Mittel gebraucht werden, sollte sich das Ethikteam umgehend an die Geschäftsführung wenden. Diese muss dann erneut in die Phase des strategischen Managements eintreten und überdenken, ob sie das betreffende Produkt wirklich produzieren sollten – ob sie dies verantworten können.

10.1 Prüfung der Kriterien für die ethische Gestaltung eines Datenökosystems

Ein Datenökosystem ist ein soziotechnisches System, das heißt, es besteht aus Technologie und sozialen Beziehungen. Um ein Datenökosystem ethisch gestalten zu können, müssen also die Technologie und die sozialen Beziehungen ethisch gestaltet werden. Durch meinen ethischen Verfahrensmaßstab wird die Technologie ethisch gestaltet, indem die bei ihrer Verwendung entstehenden Risiken explizit gemacht und minimiert werden. Gleichzeitig werden die sozialen Beziehungen im Datenökosystem ethisch gestaltet, indem die Wissens- und Kontrollverhältnisse zwischen den Akteuren herausgearbeitet und Ungleichheiten kompensiert werden. Außerdem sorgt der Erhalt eines gerechten sozialen Systems der Risikobereitschaft, von dem alle gleichermaßen profitieren, dafür, dass jegliche Risikoauferlegung durch den Einsatz von Technologie im Datenökosystem ethisch zulässig ist.

An dieser Stelle soll anhand der Kriterien aus Kapitel 9 festgestellt werden, ob sich der ethische Verfahrensmaßstab auch zur Übersetzung dieser ethischen Überlegungen in die Praxis eignet.

Allgemeine Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung

i) Die Umsetzungsmethode sollte konkrete praktische Anleitung bieten.
ii) Die Umsetzungsmethode sollte den Einfluss der Technologie auf alle Betroffenen berücksichtigen.
iii) Im Rahmen der Umsetzungsmethode sollte die Verantwortung für die Umsetzung der ethischen Überlegungen bestimmten menschlichen Akteuren zugeordnet werden.

Mein ethischer Verfahrensmaßstab ermöglicht, technische Entwicklungen ethisch zu gestalten, indem er durch die vier zu befolgenden Schritte konkrete praktische Anleitung bietet (i). Innerhalb der jeweiligen Schritte bieten Erkenntnisse aus dieser Arbeit über die Struktur von und Risiken in Datenökosystemen, sowie die Interpretation von Kants dritter Formel des Kategorischen Imperativs weitere Orientierung. So bleibt es zumindest innerhalb des Ethikteams nicht bei leeren Worten. Denn es würde bemerkt werden, wenn die konkrete praktische Anleitung nicht befolgt wird und somit für die ethischen Überlegungen keine Durchsetzungsmaßnahmen ergriffen werden. Außerdem wird durch meinen ethischen Verfahrensmaßstab der Einfluss der Technologie auf alle Betroffenen berücksichtigt (ii) – insbesondere auf die Personen, die wenig Kontrolle über die Gestaltung der Technologie und ihrer Risiken haben. Dadurch kann bei der Gestaltung der Technologie das Ziel, einen Vorteil für alle von der Technologie betroffenen Personen zu schaffen, erreicht werden. Zudem trägt das Ethikteam die Verantwortung für die Einhaltung des ethischen Verfahrensmaßstabs, und das Unternehmen trägt als kollektiver Akteur die Verantwortung für potenzielle durch die Technologie verursachte Schäden und für ihre ethische Gestaltung im Vorhinein (iii). Somit erfüllt mein ethischer Verfahrensmaßstab alle allgemeinen Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung.

Er ermöglicht es außerdem, als Umsetzungsmethode spezifisch Datenökosysteme ethisch zu gestalten.

Spezifische Kriterien für die Umsetzungsmethode zur ethischen Gestaltung eines Datenökosystems

I. Die Umsetzungsmethode sollte es ermöglichen, die Auferlegung von Risiken zu regulieren.
II. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen zu erfüllen.
III. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen.

Denn mein ethischer Verfahrensmaßstab gestattet es, die Risiken zu betrachten, welche Datengenerierende Datenquellen auferlegen und welchen Datennutzende durch Datenverarbeitende ausgesetzt sind. Weil zudem klar ist, unter welchen Umständen die Auferlegung des Risikos zulässig ist, kann diese reguliert werden (I). Laut meinem ethischen Verfahrensmaßstab dürfen Risiken nur dann auferlegt werden, wenn dadurch niemand bloß als Mittel gebraucht wird. Um dies zu realisieren, gilt das Prinzip der Maximierung der individuellen Kontrolle. Die Regulierung kann bei verschiedenen Risiken unterschiedlich ausfallen, sie unterliegt aber immer den gleichen Prinzipien. Mein ethischer Verfahrensmaßstab ermöglicht ferner die Erfüllung der administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen (II) und zur Umsetzung von Transparenz für die Datennutzenden (III). Sie dienen sogar als Prüfsteine für die Minimierung der Risiken der verletzten Privatsphäre und fehlenden Transparenz.

10.2 Die Rolle der Ethiker:innen

Da die Ethiker:innen im Ethikteam und bei der Durchführung des Verfahrensmaßstabs eine besondere Rolle spielen, soll diese hier näher beleuchtet werden. Nach Ansicht der Sozialethiker:innen Hannah Bleher und Matthias Braun bestehen bei der Ausübung der Rolle der Ethiker:innen zwei Gefahren: Zum einen könnte die ethische Entscheidungsfindung von der subjektiven Perspektive der eingebundenen Ethiker:innen abhängen, denn auch Ethiker:innen könn-

ten ihre Machtposition ausnutzen und mit ihren ethischen Überlegungen nicht nur lenken, sondern auch manipulieren.[1368] Zum anderen ist es ebenso möglich, dass die Urteile der Ethiker:innen nicht beachtet werden und damit keinen Einfluss auf den Gestaltungsprozess haben.[1369] Das liegt daran, dass ethische Erwägungen oft in Spannung stehen zu anderen Zielen der Unternehmen. Insbesondere weil diejenigen, die die anderen Ziele der Unternehmen verfolgen, mehr Macht haben als die eingebundenen Ethiker:innen besteht die Gefahr, dass die ethischen Erwägungen ignoriert werden.

Damit die Berücksichtigung ethischer Überlegungen bei der Gestaltung eines Datenökosystems nicht nur ein leeres Versprechen bleibt, also das Ethikteam nicht ignoriert wird und somit keinen Einfluss auf den Gestaltungsprozess bekommt[1370], sollte es laut McLennan et al. Durchsetzungsmaßnahmen geben, wie zum Beispiel Regulierung, Zertifizierung oder freiwillige Maßnahmen. Sie empfehlen zusätzlich, dass der ganze Prozess beaufsichtigt werden sollte und dass für diese Aufsicht Richtlinien erstellt werden sollten.[1371] Ich argumentiere, dass die am Ethikteam beteiligten Mitarbeitenden als eine interne Aufsicht fungieren. Denn sie lernen in Treffen mit dem Ethikteam die Relevanz ethischer Überlegungen kennen und beobachten dann auf ihren Unternehmensebenen, ob diese Überlegungen auch wirklich in die Gestaltung miteinfließen. Wenn hier Diskrepanzen entstehen, sollten diese im Ethikteam gemeldet und anschließend mit den für die Umsetzung der ethischen Anforderungen Verantwortlichen diskutiert werden. Lässt sich die Uneinigkeit auf diese Weise nicht klären, so gilt die Geschäftsführung als Garant für die Umsetzung. Sie ist es, die beschlossen hat, das entsprechende Produkt herzustellen, und sie sollte ein Interesse daran haben, dass dieses Produkt ethisch gestaltet wird. Ist dieses Interesse nicht vorhanden, gibt es wahrscheinlich gar kein Ethikteam, denn seine Aufstellung muss durch die Geschäftsführung veranlasst werden.[1372]

1368 Bleher, Braun 2023:7
1369 Ebd.
1370 Ebd.:7
1371 McLennan et al. 2020:489; McLennan et al. 2022:8
1372 Dazu mehr in Unterkapitel 10.5

Die Rolle der Ethiker:innen ist anspruchsvoll. Sie müssen ihr Team anleiten, Probleme antizipieren, sich im Unternehmen durchsetzen, dabei immer begründen, was sie tun und vor allem plausibel machen, warum mit den ethischen Problemen umgegangen werden sollte. In ihrer Rolle praktizieren sie Angewandte Ethik. Das heißt, die Ethiker:innen vermitteln zwischen den Grundlagen der Ethik und konkreten Einzelfällen, die auch durch empirische Tatsachen und spezifische Gegebenheiten geprägt sind. Damit ist die Angewandte Ethik ein Teilbereich der Praktischen Ethik, welche sich mit ethischen Fragen befasst, die sich im Rahmen bestimmter sozialer Praktiken stellen.[1373] Diese umfassende Perspektive der Praktischen Ethik auf die Anwendung ethischer Prinzipien im Alltag bildet also den Hintergrund, aber die Ethiker:innen, welche die ethische Gestaltung von Datenökosystemen anleiten sollen, konzentrieren sich auf spezifische Anwendungsbereiche – in diesem Fall auf Technologieethik. Das Ziel ist es, in den Anwendungsbereichen ethische Reflexion zu fördern und in Konfliktsituationen Entscheidungshilfen zu entwickeln.

Deshalb müssen die Ethiker:innen auch über Wissen in diesem Anwendungsbereich verfügen und zum Beispiel Technologien der Datenverarbeitung oder Speicherprozesse verstehen. Bei konkreten technischen Fragen arbeiten sie, gemäß meinem Vorschlag eines Ethikteams, mit anderen Teammitgliedern zusammen, die beispielsweise auf der Ebene der Entwicker:innen beschäftigt sind. Somit erfordert diese Rolle als Ethiker:in auch ein gewisses Selbstverständnis. Die Ethiker:innen sollten sich als eine Art Vermittlung verstehen zwischen ethischen Überlegungen und konkreten Anwendungsfällen. Dabei sollte weder der präferierte ethische Ansatz noch der technische Status quo unhinterfragt gesetzt werden. Insbesondere wenn technische Pläne mit den ethischen Vorstellungen kollidieren, sind Kreativität und eine Offenheit gegenüber neuen Lösungsmöglichkeiten gefragt. Hier ist es die Rolle der Ethiker:innen diese gemeinsame Ideenfindung im Team anzuleiten. Aber auch wenn Zweifel geäußert werden an den ethischen Voraussetzungen, ist es die Aufgabe der Ethiker:innen, offen zu sein für konstruktive Kritik. Mein ethischer Verfahrensmaßstab kann an dieser Stelle als klare Anleitung der ethischen Voraussetzungen dienen. Er gilt nicht un-

1373 Lindemann 2019:4f.

hinterfragt, wurde aber mit dem deontologischen Anspruch an ein wünschenswertes gesellschaftliches Miteinander logisch hergeleitet.

10.3 Herleitung des ethischen Verfahrensmaßstabs

Der ethische Verfahrensmaßstab kann aus den ethischen Überlegungen in dieser Arbeit hergeleitet werden. Denn die gesamte bisherige Analyse des Umgangs mit digitalen Daten im Datenökosystem und den dabei entstehenden Risiken mündet in diesem Vorschlag zur ethischen Gestaltung des Ganzen. Die Annahmen über ein wünschenswertes gesellschaftliches Miteinander, die diesem Maßstab zugrunde liegen, begründen sich in Kants erster und dritter Formulierung des Kategorischen Imperativs. Sie bauen somit auf ein in sich stimmiges ethisches System auf. Im Folgenden soll gezeigt werden, wie sich die einzelnen Schritte meines ethischen Verfahrensmaßstabs aus der bisherigen Arbeit herleiten lassen.

In Anlehnung an das interne Audit aus Kapitel 9.3.2 wird der ethische Verfahrensmaßstab intern durchgeführt.[1374] Dabei soll allerdings in Anlehnung an Hanssons Risikoanalyse aus 9.3.3 eine Outsider-Perspektive eingenommen werden.[1375] Im Prinzip geht es wie bei der Responsible Innovation Methode aus Kapitel 9.2.3 darum, Risiken zu antizipieren, die eigene Rolle zu reflektieren und darüber zu deliberieren, sowie responsiv mit Lösungsvorschlägen darauf einzugehen.[1376] Bei diesem ethischen Verfahrensmaßstab wird dabei jedoch mit Hilfe von Hansson ein Schwerpunkt auf die Verteilung von Risiko, Entscheidungsmacht und Begünstigung gelegt.[1377] Außerdem wird mit Hilfe der in dieser Arbeit entwickelten deontologischen Risikoethik ein Maßstab dafür angelegt, wann die Risikoauferlegung zulässig ist.

Die Verantwortung, die ich dem Unternehmen für die ethische Gestaltung seiner Produkte und dem Ethikteam für die Durchführung des Verfahrensmaßstabs zuschreibe, begründen sich in ihrer Möglichkeit zur Kontrolle über Gestaltung und Durchführung.

1374 Raji et al. 2020:33f.
1375 Ebd.:1821f.
1376 Ebd.:44
1377 Hansson 2018:1822

Denn wenn sich Risikoexponierte nicht freiwillig dem Risiko aussetzen, sondern sie durch bestimmte Umstände dazu gezwungen sind, liegt laut Hermansson und Hansson die Verantwortung für die Vermeidung des Schadeneintritts und die Erreichung von guten Ergebnissen bei denjenigen, die das Risiko und den potenziellen Schaden verursachen.[1378] Bei der Gestaltung eines Datenökosystems muss außerdem jemand vorausschauend Verantwortung übernehmen für eine möglichst gute Gestaltung des Systems, weil das Fehlen von verantwortungsvollem Handeln zu Schaden führen kann und es im Fall eines Schadens wichtig ist, dass jemand dafür zur Rechenschaft gezogen werden kann.[1379]

Bei der Übernahme der vorausschauenden Verantwortung muss das Unternehmen Unsicherheiten über die Zukunft berücksichtigen und mit Fürsorge schon vor der Entwicklung neuer Technologien darüber nachdenken, was diese (nicht) tun sollen, sowie während der Entwicklung auf die Ansichten, Perspektiven und Rahmenbedingungen aller Interessengruppen reaktionsfähig einzugehen.[1380] Dabei geht es einerseits um die Vermeidung von zukünftigen Schäden (negative vorausschauende Verantwortung) und andererseits darum, gute Ergebnisse herbeizuführen (positive vorausschauende Verantwortung).[1381] Wenn das Unternehmen als kollektiver Akteur mit Kontrolle weiß, dass es zur Verantwortung gezogen werden kann, führt das zu bedachtem Verhalten auf Seiten der Handelnden, und sorgt für ein Gefühl der Sicherheit auf Seiten der Behandelten.[1382] So geht es im Datenökosystem bei der Übernahme von Verantwortung darum, dass Datenquellen und Datennutzende sicher sein können, dass sich jemand um den Schutz ihrer Privatsphäre beziehungsweise um die Gewährleistung von Transparenz kümmert.

1378 Hermansson, Hansson 2007:142
1379 Danaher 2016:299; Kiener 2022:584; Nyholm 2024:192; Oimann, Tollon 2024:5
1380 Owen et al. 2013:29, 35f.
1381 Nyholm 2024:193, 199
1382 Young 2011:105f.

Schritt 1: Positionierung zu Verantwortung, Nutzen und Risiken

Dadurch, dass sich das Ethikteam beim Durchführen des ethischen Verfahrensmaßstabs als erstes bewusst macht, wer verantwortlich ist für die ethische Gestaltung der betreffenden Technologien bzw. des gesamten Datenökosystems, wird es sich seiner Rolle bewusst. Denn es tragen diejenigen Verantwortung dafür, dass niemandem unzulässigerweise ein Risiko auferlegt wird, dass Datenquellen nicht die Kontrolle über ihre Informationen verlieren und dass Datennutzende nicht durch Verzerrungen und Intransparenz zu falschen Überzeugungen gelangen, die entweder Kontrolle über die betreffenden Handlungen und ihre absehbaren Folgen haben[1383] oder bei unabsehbaren Folgen eine beaufsichtigende Rolle im Sinne Nyholms[1384] einnehmen. Da es die Datengenerierenden sind, die kontrollieren können, wie viel Kontrolle die Datenquellen über ihre Informationen haben, sind sie dafür verantwortlich, die Datensammlung und -speicherung so zu gestalten, dass Datenquellen möglichst viel Kontrolle haben und möglichst wenig Autonomie verlieren. Da die Datenverarbeitenden eine gewisse Kontrolle haben über die Transparenz der von ihnen bereitgestellten Ergebnissen und mindestens eine beaufsichtigende Rolle einnehmen, sind sie dafür verantwortlich die Verarbeitung der Daten und das Bereitstellen von Ergebnissen möglichst transparent zu gestalten und möglichst keine falschen Überzeugungen zu verursachen.

In der Realität bestehen die Akteure Datengenerierende und Datenverarbeitende aus vielen verschiedenen Mitarbeitenden aus allen drei Entscheidungsebenen, die es laut Gogoll et al. bei der Software-Entwicklung in einem Unternehmen gibt: Strategisches Management, Konzeptentwicklung, Entwicklungsprozess.[1385] Das heißt, es handelt sich um kollektive Akteure. Dazu kommt, dass es ein Unternehmen sein kann, welches die Rollen von sowohl Datengenerierenden als auch Datenverarbeitenden sowie Verwaltenden und Infrastrukturbereitstellenden einnimmt – insbesondere, wenn ein Unternehmen mit der Gestaltung eines Datenökosystems beauftragt wurde oder aus eigenen Interessen ein Datenökosystem gestalten

1383 Frankfurt 1971; Fischer, Ravizza 1998; Kane 1996; Smith 2005
1384 Nyholm 2018:1216
1385 Gogoll et al. 2021:1087

möchte. Deshalb fällt in meinem ethischen Verfahrensmaßstab die Verantwortung für die ethische Gestaltung des gesamten Datenökosystems bzw. jeder der eingesetzten Technologien dem gesamten Unternehmen als kollektivem Akteur zu. Dabei wird angenommen, dass besagtes Unternehmen auf allen Entscheidungsebenen die Rollen der Datengenerierenden und der Datenverarbeitenden einnimmt und dadurch Kontrolle ausübt über die Gestaltung[1386] oder zumindest eine Beaufsichtigungsrolle[1387] ausübt.

Da innerhalb des Unternehmens spezifisch dem Ethikteam die Kontrolle über die Umsetzung des ethischen Verfahrensmaßstabs zugeteilt wird, ist es das Ethikteam, welches die Verantwortung für die Einhaltung des ethischen Verfahrensmaßstabs trägt. Doch es ist das gesamte Unternehmen als kollektiver Akteur, welches die finale Gestaltung des Datenökosystems verantworten muss. Sollten Risiken unzulässigerweise auferlegt werden, sollten Datenquellen die Kontrolle über ihre Informationen verlieren oder sollten Datennutzende durch Verzerrungen und Intransparenz zu falschen Überzeugungen gelangen, muss also das Unternehmen dafür geradestehen. Diese moralische Verantwortung soll ein Anreiz sein für bedachtes Verhalten auf Seiten der gestaltenden Unternehmen und sie soll bei Datenquellen und Datennutzenden ein Gefühl der Sicherheit auslösen.[1388]

Erst wenn man sich damit befasst, welchen Nutzen Datenverarbeitung im Kontext von Datenökosystemen hat, kann man abwägen, ob es sich auch lohnt, die Risiken dafür in Kauf zu nehmen. Diese Risiken entstehen nicht erst bei der Anwendung des fertig entwickelten Systems, sondern sie können schon allein dadurch entstehen, dass Vorurteile unreflektiert in die zu gestaltende Technologie miteinfließen (preexisting bias) oder dass technische Beschränkungen zu einer systematischen Ungleichbehandlung führen können (technical bias). Außerdem kann der Kontext, in dem die zu gestaltende Technologie zukünftig eingesetzt wird, ein Risiko bergen (emergent bias). Durch die Einstufung dieser möglichen Verzerrungen als Risiken, wurde in dieser Arbeit gezeigt, warum es wichtig ist, sich mit den Risiken im Vorhinein auseinanderzusetzen. Denn alle Verzerrungen können durch besondere Aufmerksamkeit minimiert

1386 Frankfurt 1971; Fischer, Ravizza 1998; Kane 1996; Smith 2005
1387 Nyholm 2018:1210; Nyholm 2024:204
1388 Young 2011:105f.

oder sogar vermieden werden. So können in Schritt 1 Vorurteile reflektiert werden, und es kann die Nutzung des Systems antizipiert werden, um die Auswirkungen von technischen Entscheidungen und unpassenden Nutzungskontexten im Vorhinein zu betrachten.[1389]

Bei der Anwendung des fertig gestalteten – im besten Fall verzerrungsfreien – Systems entstehen trotzdem weitere Risiken, die in dieser Arbeit ausführlich betrachtet wurden: die Verletzung der Privatsphäre von Datenquellen und das Fehlen von Transparenz für Datennutzende. Zum einen können beim Eintritt dieser Risiken unerwünschte Handlungsfolgen entstehen (Kapitel 6 und 7), zum anderen verursacht allein schon die Risikoauferlegung einen Schaden (Kapitel 5). Das heißt, der Nutzen des zu entwickelnden Datenökosystems, bzw. einer seiner Technologien, muss groß sein, um all diese Risiken rechtfertigen zu können.

Die explizite Betrachtung von Nutzen und Risiken der zu gestaltenden Technologie ist auch deshalb ein wichtiger erster Schritt, weil er die Grundlage schafft für die nächsten Schritte. Denn erst durch die Konzentration auf Nutzen und Risiko fällt einem auf, wer den Nutzen hat und wer die Risiken trägt. So werden die Kontrollungleichheiten zwischen den Akteuren deutlich, die für die folgenden Schritte ausschlaggebend sind.

Schritt 2: Risikoverteilung bewerten mit Hanssons ethischer Risikoanalyse

Hanssons ethische Risikoanalyse (eRA) ermöglicht es, die zwischenmenschlichen Beziehungen der von der Entwicklung eines Datenökosystems Betroffenen in den Vordergrund zu rücken. Dies ist so wichtig, weil in dieser Arbeit gezeigt wurde, dass die Akteure im Datenökosystem unterschiedliche Kontrollmöglichkeiten haben und es somit zu schädlicher Risikoauferlegung kommen kann. Diese Risikoauferlegung ist möglich durch eine schlechtere epistemische Situation für Datenquellen und Datennutzende. Denn den Datenquellen fehlt allein durch ihre Rolle im Datenökosystem Wissen über den Umgang mit ihren Daten. Bei der Sammlung ihrer Daten können sie die Kontrolle über diese verlieren, sodass ihnen durch Datengenerie-

1389 Friedman, Nissenbaum 1996:343f.; Friedman, Kahn 2008:1253

rende das Risiko der Verletzung ihrer Privatsphäre auferlegt wird. Datennutzende werden von Datenverarbeitenden in eine Situation gebracht, in welcher sie den ihnen zu Verfügung gestellten Ergebnissen einfach glauben müssen. Dabei können sie auch falschen Ergebnissen ausgesetzt sein, ohne dies überprüfen zu können – ihnen wird somit durch fehlende Transparenz das Risiko auferlegt, zu falschen Überzeugungen zu gelangen.

Diese zwischenmenschlichen Beziehungen im Datenökosystem können als epistemische Asymmetrien bezeichnet werden, weil zwei der Akteure jeweils nicht über das Wissen der anderen beiden Akteure verfügen.[1390] In Datenökosystemen herrscht eine epistemische Asymmetrie, weil Datengenerierende und Datenverarbeitende Wissen aus erster Hand über die Sammlung, Speicherung und Analyse von Daten besitzen. Datenquellen und Datennutzende hingegen sind auf sekundäres Wissen angewiesen. Diese ungleiche Wissensverteilung verleiht Datengenerierenden und Datenverarbeitenden eine epistemische Autorität, die sie auch strategisch einsetzen könnten – was jedoch nicht zwangsläufig geschieht.[1391]

Von epistemischer Dominanz spricht man, wenn ein Akteur kontrolliert, welche Beweise einem anderen zur Verfügung gestellt werden und dadurch dessen Überzeugungen beeinflussen kann.[1392] Datengenerierende können Informationen über Datennutzung kontrollieren, während Datenverarbeitende die Nachvollziehbarkeit algorithmischer Ergebnisse beeinflussen. Dennoch sind Datenquellen und Datennutzende nicht vollständig abhängig – sie können sich zusätzlich aus externen Quellen informieren. In konkreten Systemen jedoch bleibt entscheidendes Wissen bei den dominierenden Akteuren. Diese Dominanz muss nicht negativ sein: Bei guter Absicht kann sie sogar epistemisch nützlich sein. Problematisch wird sie aber bei Fehlinformation oder bewusster Irreführung – etwa wenn Datensicherheit beschönigt oder algorithmische Verzerrungen verschleiert werden. Doch selbst bei bester Absicht bleibt eine epistemische Unsicherheit bestehen, da Risiken wie Datendiebstahl oder algorithmische Fehler nie ganz ausgeschlossen werden können.[1393]

1390 Drew 1991:22, 25; Russo et al. 2024:1592
1391 Drew 1991:32, 39f.
1392 Harris 2022:134, 137
1393 Möller 2009:219f.

Ein zusätzlicher Konflikt entsteht, wenn die Akteure, die Risiken einschätzen, auch diejenigen sind, die diese Risiken anderen auferlegen. In solchen Fällen könnten Interessen eine Rolle spielen, die eine objektive Einschätzung erschweren. Am Ende tragen Datenquellen das Risiko von Datenschutzverletzungen und Datennutzende das Risiko, durch intransparente oder fehlerhafte Ergebnisse zu falschen Überzeugungen zu gelangen.

Genau diese Problematik kann mit Hilfe von Hanssons ethischer Risikoanalyse systematisch abgebildet werden. Denn das zugrundeliegende Drei-Parteien Modell von Hermansson und Hansson (Kapitel 3.2.1) ermöglicht ein besseres Verständnis der Beziehung zwischen den Akteuren im Datenökosystem, indem bei der Risikoauferlegung drei relevante Parteien identifiziert werden: die Risikoexponierten, die Entscheidungstragenden und die durch das Risiko Begünstigten.[1394] Hierdurch wird klar, dass neben der epistemischen Asymmetrie auch asymmetrische Kontrollmöglichkeiten vorliegen können. Denn im zweiten Schritt der eRA werden die Rollenkombinationen und Wechselbeziehungen zwischen den Parteien betrachtet, um feststellen zu können, ob die Risikoexponierten über das Risiko mitentscheiden können und ob sie auch durch das Risiko begünstigt werden. Wenn Nutzen und Risiken zwischen den Akteuren ungleich verteilt sind, wird die Rollenverteilung durch eine Verteilungsanalyse bewertet. Ist die Rollenverteilung ungerecht auf Grund von ungleichen Kontrollmöglichkeiten, werden die zugrunde liegenden Machtstrukturen analysiert.[1395]

Die eRA ermöglicht es festzustellen, ob ungleiche Kontrollmöglichkeiten vorherrschen und somit eine Risikoauferlegung vorliegt. Ob aber diese Risikoauferlegung zulässig ist, wird erst im nächsten Schritt meines ethischen Verfahrensmaßstabs analysiert.

Schritt 3: Überprüfung der Zulässigkeit von Risikoauferlegung

Die Auferlegung eines Risikos ist deshalb moralisch signifikant, weil sie mit dem Verlust der Autonomie für die Risikoexponierten einhergeht. Wem ein Risiko auferlegt wird, der hat nicht die Wahl, ob

1394 Hermansson, Hansson 2007:130
1395 Hansson 2018:1820ff.

er ein Risiko eingehen oder vermeiden will – mit den Handlungsoptionen ist dann auch seine Autonomie beschränkt. Denn Risikoexponierte werden durch eine Risikoauferlegung fremdbestimmt und können in diesem Kontext nicht mehr so leben, wie sie wollen.[1396] Die Autonomie der Risikoexponierten sollte jedoch nicht auf diese Weise beschränkt werden, da Autonomie eine notwendige Bedingung für ein gelungenes Leben ist.[1397] Hinzu kommt, dass die Auferlegung von Risiken durch die aufgezwungene Unsicherheit über mögliche negative Folgen psychischen Stress verursacht.[1398] Schon die Möglichkeit eines Schadens wird als belastend empfunden, sodass manche Menschen sogar einen sicheren Schaden der Ungewissheit vorziehen.[1399]

In Kapitel 5 wurde die Pflicht begründet, durch das eigene Handeln andere keinem Risiko auf Schaden gegen ihren Willen und ohne zureichende Aufklärung auszusetzen. Da sich das Auferlegen von Risiken aber in einer Gesellschaft nicht insgesamt vermeiden lässt, muss es Bedingungen geben, unter denen diese Pflicht verletzt werden darf. Hansson meint, eine Person darf dann und nur dann einem Risiko ausgesetzt werden, wenn dieses Aussetzen Teil eines gerechten sozialen Systems der Risikobereitschaft ist, von dem sie selbst profitiert.[1400] Da ich aber der Ansicht bin, dass es für gerechtes soziales System der Risikobereitschaft nicht reicht, wenn ein Risiko durch einen eigenen Vorteil kompensiert wird, führe ich in Kapitel 5 Kants dritte Formel des Kategorischen Imperativs als Regel für ein solches System ein und begründe eine deontologische Risikoethik. Demnach sollten Risikoexponierte über das ihnen auferlegte Risiko informiert werden und sie sollten idealerweise die Wahl haben, ob sie das Risiko tragen wollen. Wenn dies nicht umsetzbar ist, sollte das Risiko so gering wie möglich sein, und Risikoexponierte sollten für das Risiko sowie für mögliche Schäden Kompensation erhalten.[1401] Außerdem sollte bei jeder Risikoauferlegung das Prinzip der Maximierung der individuellen Kontrolle gelten, damit die Risikoexponierten den größtmöglichen Entscheidungsspielraum haben

1396 Rössler 2001:39, 103, 119; Oberdiek 2017:90
1397 Rössler 2001:96,126
1398 Wolff, De-Shalit 2007:68
1399 Luhmann 1990:159; Wolff 2021:12; Roeser 2014:640, 648
1400 Hansson 2003:305
1401 Ebd.:303

und sie das Risiko mit Hilfe von relevanten Informationen ohne Zwang und Täuschung verstehen können.

In Schritt drei meines ethischen Verfahrensmaßstabs wird also die Zulässigkeit der Risikoauferlegung geprüft, weil die Pflicht besteht, anderen ohne deren Aufklärung und Kontrolle kein Risiko aufzuerlegen, da dies zu Autonomieverlust und weiteren negativen Folgen wie psychischem Stress führen kann. Um ein Leben in der Gesellschaft, in welcher sich Menschen zwangsläufig gegenseitig Risiken aussetzen, zu ermöglichen, habe ich in dieser Arbeit Bedingungen begründet, unter denen dies ethisch zulässig ist. Angenommen, alle möchten in einer Gesellschaft leben, in der niemand als bloßes Mittel gebraucht wird, dann sollte bei einer Risikoauferlegung auch niemand als bloßes Mittel gebraucht werden. Das heißt, niemand sollte anderen ein Risiko auferlegen, ohne ihnen die Wahl zu lassen und ohne sie darüber zu informieren, was sie wählen.

In diesem dritten Schritt muss also geprüft werden, ob die Risikoexponierten die Wahl haben, das Risiko zu akzeptieren oder abzulehnen und ob sie über relevante Informationen verfügen, um diese Entscheidung informiert treffen zu können. Notwendig für ihre Zustimmung ist, dass der Zweck des Risikos die Achtung der Menschlichkeit fördert. Hilfreich für die Zustimmung wäre es, wenn die Risikoexponierten durch die Risikoauferlegung auch darin befördert würden, ihre eigenen qualifizierten Zwecke zu verfolgen und wenn sie sogar selbst von dem Risiko profitierten. Ist den Risikoexponierten die Wahl verwehrt, steht ihnen Kompensation für mögliche Schäden zu.[1402]

Schritt 4: Fehlervorsorge durch Minimierung der Risiken

Neben dem möglichen Mangel an Autonomie sowie der aufgezwungenen Unsicherheit, welche selbst als Schaden empfunden werden kann, wird bei jeder Risikoauferlegung auch der Eintritt eines Schadens riskiert. Deshalb geht es in Schritt 4 des ethischen Verfahrensmaßstabs um die potenziellen Folgen bei einem Schadenseintritt, die laut dem Kategorischen Imperativ vermieden werden sollten.

1402 Hansson 2003:303

In dieser Arbeit wurde gezeigt, dass es sich bei den Risiken im Datenökosystem zum einen um die potenzielle Verletzung der Privatsphäre der Datenquellen handelt und zum anderen um die potenziell fehlende Transparenz für die Datennutzenden. Die moralische Pflicht der Datengenerierenden und Datenverarbeitenden zum Schutz der Privatsphäre der Datenquellen begründet sich dabei sowohl in den negativen Konsequenzen, die eine Verletzung der Privatsphäre verursachen kann – diese kann Erpressung, staatliche Kontrolle oder kommerzielle Ausbeutung zur Folge haben[1403] – als auch in den positiven Konsequenzen, die eine intakte Privatsphäre haben kann. Denn nur bei einer intakten Privatsphäre können sich Datenquellen autonom entfalten und können zum Beispiel intime Beziehungen der Liebe und der Freundschaft eingehen – das fördert ihr Wohlbefinden, sodass sie ihrem Selbstzweck gerecht werden und ihre positive Freiheit ausleben können.[1404]

Zum Bereitstellen von transparenten algorithmischen Ergebnissen besteht eine Pflicht für Datengenerierende und Datenverarbeitende, weil Datennutzende epistemisch abhängig sind von den Ergebnissen und auf dieser Informationsgrundlage falsche Überzeugungen bilden können. Nur mit Hilfe der transparenten Darstellung können Datennutzende eine fundierte Entscheidung darüber treffen, wie sie mit den betreffenden Ergebnissen umgehen[1405] und können auf Grundlage dieser Informationen ihre eigenen Interessen vertreten[1406]. Datengenerierende und Datenverarbeitende sollten auf diese Weise dafür sorgen, dass möglichst keine falschen Überzeugungen entstehen, weil es für das menschliche Wohlergehen wichtig ist, dass wir uns von richtigen und nicht von falschen Annahmen leiten lassen.[1407] Außerdem lässt sich die Pflicht zur Transparenz dadurch begründen, dass fehlende Transparenz dazu führen kann, dass bestimmte Gruppen auf Grund von Merkmalen benachteiligt werden, die eigentlich keine Rolle spielen sollten.[1408] Eine solche Diskriminierung bestimmter Gesellschaftsgruppen bleibt mit größe-

1403 Moore 2010:97 Rössler 2003:3
1404 Masur et al. 2018:6; Westin 1967, S. 35–43; Fried 1968:216; Berlin 1997:196, 202f.
1405 Munch et al. 2024:9
1406 Vredenburgh 2022:212
1407 Alston 2005:30
1408 Stahl et al. 2023:10f.

rer Wahrscheinlichkeit unbemerkt, wenn die Entstehung der Ergebnisse intransparent ist, und das kann nicht gewollt werden.[1409]

Aus der Definition von Privatsphäre als Wahl-Kontrolle des Individuums über Zugang zu und Benutzung von seinen persönlichen Informationen lassen sich Anforderungen an den Umgang mit Daten und insbesondere mit den Datenquellen im Datenökosystem ableiten. Da Wahl-Kontrolle bedeutet, dass man entscheiden kann, mit wem man welche Informationen zu welchem Zweck teilt und eine Pflicht zu deren Schutz besteht, sollten die betroffenen Personen also zum einen entscheiden können, wer Zugriff auf welche Informationen bekommt. Zum anderen sollten die betroffenen Individuen entscheiden können, zu welchem Zweck ihre Informationen gesammelt oder verarbeitet werden. Administrativ bedeutet das, dass einerseits das Einverständnis der Individuen eingeholt werden sollte, und die Individuen andererseits über jeden Verwendungszweck der Daten informiert sowie die Daten vor Zweckentfremdung geschützt werden sollten.

Aus den theoretischen Überlegungen zur Transparenz als Zugänglichkeit und Verständlichkeit[1410] algorithmischer Ergebnisse lassen sich Anforderungen an den Umgang mit Daten bei der Datenverarbeitung ableiten. Um den Datennutzenden nicht das Risiko der fehlenden Transparenz im Rahmen eines gerechten sozialen Systems der Risikobereitschaft aufzuerlegen, sollten Datenverarbeitende ihrer Pflicht zur Transparenz nachkommen, indem sie die algorithmischen Ergebnisse im zum Verständnis notwendigen Kontext darstellen. Nur so können die Datennutzenden nachvollziehen, wie das Ergebnis zu verstehen ist und dementsprechend ihre eigenen Interessen vertreten. Administrativ bedeutet das, dass die Datenverarbeitenden dazu in der Lage sein müssen nachzuvollziehen, wie das Ergebnis zustande gekommen ist bzw. welcher Input zu welchem Output führt und welche Trainingsdaten für die eingesetzte KI verwendet wurden.

Nur wenn diese administrativen Anforderungen erfüllt sind, werden Privatsphäre und Transparenz sowohl geschützt als auch aktiv gefördert. Deshalb werden in Schritt 4 des ethischen Verfahrensmaßstabs die administrativen Anforderungen umgesetzt mit Hilfe

1409 Hare 1993:15
1410 Mittelstadt et al. 2016:6

von technischen und kommunikativen Maßnahmen. Denn die Umsetzung der administrativen Anforderungen bedeutet eine Minimierung der Risiken. Dies ist, wie in dieser Arbeit gezeigt wurde, ethisch zwingend notwendig, um die Autonomie der Datenquellen sowie die Fähigkeit zur Selbstvertretung durch die Datennutzenden zu schützen und zu fördern.

Wenn die Umsetzung der administrativen Anforderungen und damit das Minimieren der Risiken nicht möglich ist und die Risikoexponierten dadurch nicht über relevante Informationen verfügen, um eine informierte Entscheidung treffen zu können, ob sie das Risiko tragen oder ablehnen wollen, sodass die Rollenverteilung ungerecht bleibt auf Grund von ungleichen Kontrollmöglichkeiten, sollte die Geschäftsführung überdenken, ob sie verantworten kann, das betreffende Produkt zu produzieren.

Die Umsetzung der Ethik in die Praxis mit Hilfe meines ethischen Verfahrensmaßstabs soll den Handlungsspielraum der Unternehmen erweitern, blinde Flecken aufdecken, Eigenverantwortung stärken und die Autonomie von allen Betroffenen fördern – so wie Hagendorff das von der Ethik erwartet.[1411] Somit soll die Ethik in Form des Verfahrensmaßstabs dazu beitragen, die Aktivitäten der Unternehmen zu fördern oder sogar zu verbessern.[1412] Damit diese Methoden erfolgreich in der Praxis eingeführt werden kann, werden keine erheblichen Änderungen der bestehenden Prozesse gefordert, sondern es wird ein Gleichgewicht zwischen Strenge und einfacher, relativ nahtloser Umsetzung angestrebt.[1413]

10.4 Antwort auf mögliche Kritik an meinem ethischen Verfahrensmaßstab

Es ist vorstellbar, dass die rein theoretische Herleitung meiner ethischen Anforderungen kritisiert wird. Man könnte argumentieren, dass man durch das Vorschlagen vordefinierter ethischer Prinzipien Gefahr läuft, die unterschiedlichen Werte und Bedürfnisse verschie-

1411 Hagendorff 2020:112f.
1412 Boddington 2017:8
1413 Murphy 2023:4

dener Nutzergruppen zu übersehen.[1414] Deshalb wird in vielen der Gestaltungsprozess-Ansätzen großer Wert auf die Integration der Stakeholder gelegt. Der Schwerpunkt des Human-centered Design hat gezeigt, dass durch die Integration von Stakeholdern in den Gestaltungsprozess kontextspezifische Lösungen gefunden werden können für die Gruppe der befragten Stakeholder.[1415] Auch im Kontext des Responsible Innovation Ansatzes ermöglicht die Integration von Stakeholdern einen deliberativen Gestaltungsprozess. Das heißt, es gibt einen offenen, strukturierten Austausch zwischen von der Technologie Betroffenen auf Basis von genauen und transparenten Informationen. Dadurch werden die Interessen, Perspektiven, Visionen, Fragen und Probleme der Betroffenen in Bezug auf das Produkt berücksichtigt.[1416]

Das Problem dabei ist, dass sich die von wenigen Betroffenen erarbeiteten Lösungen nur schwer auf größere Populationen übertragen lassen.[1417] Dazu kommt noch, dass Stakeholder in der Praxis meist erst dann eingebunden werden, wenn die wichtigsten Entscheidungen bereits getroffen sind, sodass ihre Interessen nur eingeschränkt berücksichtigt werden können.[1418] Damit Stakeholder rechtzeitig integriert werden könnten, müsste es einen methodischen Ansatz zur Identifizierung der Stakeholder geben[1419] und eine klare Anleitung für die Integration der Stakeholder[1420]. Selbst dann ließe es sich nicht vermeiden, dass Machtungleichgewicht und Interessenkonflikte eine Rolle spielen.[1421] Denn welche Parteien (an den Tisch) eingeladen werden, hängt laut der interdisziplinären Technikforscherin Carolyn Ten Holter mit Machtfragen zusammen. Diese Macht ist mit der Fähigkeit verbunden, eine Einladung aussprechen oder verweigern zu können.[1422] Aber selbst wenn alle relevanten Stakeholder an einem Tisch sitzen, sind noch Machtungleichgewichte vorhanden. Denn wenn sich nicht alle Stakeholder einig sind,

1414 Kellmeyer 2024:252
1415 Ebd.:257f.
1416 Owen et al. 2013:38; Niehaves, Heger 2015:172; Lubberink et al. 2017:15
1417 Kellmeyer 2014:257f.
1418 Stahl et al. 2021:189; Silva et al. 2019:16
1419 Winkler, Spiekermann 2018:19f.
1420 Oetzel, Spiekermann 2014:133f.
1421 Silva et al. 2019:16
1422 Holter 2022:280

manche aber mehr Einfluss haben als andere, ist es möglich, dass eigennützige Interessen mächtigerer Gruppen eine größere Rolle spielen als die Einbeziehung gesellschaftlicher und ethischer Aspekte.[1423] Deshalb schlägt Hansson vor, nicht unbedingt einen Konsens zwischen den Beteiligten anzustreben, sondern auf die unterschiedliche Verhandlungsmacht der verschiedenen Stakeholder durch die Analyse der Unterschiede in Bezug auf Risikoexposition, Macht und Ressourcen aufmerksam zu machen – das, was ich mit Hilfe von Hanssons ethischer Risikoanalyse auch in meinem Verfahrensmaßstab umgesetzt habe.[1424]

Mein Lösungsvorschlag für die Beseitigung der Umsetzungsschwierigkeiten bei der Stakeholder-Integration ist, dass die Stakeholder nicht direkt befragt, sondern konzeptionell einbezogen werden. Anstatt nach der Insider-Perspektive zu fragen, nimmt mein ethischer Verfahrensmaßstab mit Hilfe von Hanssons ethischer Risikoanalyse eine unparteiische Outsider-Perspektive ein.[1425] So wird rein theoretisch ermittelt, wer alles von der zu bewertenden Technologie betroffen sein könnte und in welcher Rolle. Es ist schwer nachzuvollziehen, ob die Deliberation mit den Stakeholdern in der Praxis überhaupt direkte Auswirkungen auf die Innovation hat, und ob die Ergebnisse des Innovationsprozesses den Wünschen der Stakeholder entsprechen.[1426] Aber wenn die Interessen der Stakeholder von Anfang an konzeptionell einbezogen werden, haben diese definitiv Auswirkungen auf die zu entwickelnde Technologie.

So wie die verschiedenen Werte von Ethischen Leitlinien miteinander in Konflikt geraten können[1427], ist es auch möglich, dass die verschiedenen Risiken sich widersprechende Maßnahmen verlangen. So könnte es vorkommen, dass beim Schutz der Privatsphäre Abstriche gemacht werden müssten, um die Transparenz der Ergebnisse zu verbessern. Denn wenn möglichst viele Informationen über die analysierten Daten zur Verfügung stehen, ist transparenter, wie die Ergebnisse zustande gekommen sind – es lassen sich womöglich sogar

1423 Blok, Lemmens 2015:23
1424 Hansson 2018:1821f.
1425 Ebd.
1426 Stahl et al. 2021:180f.
1427 Jobin et al. 2019:396

leichter Verzerrungen in den Daten erkennen, welche im Ergebnis reproduziert wurden. Das Ethikteam muss solche trade-offs von Fall zu Fall abwägen. Wenn es zum Beispiel darum geht, das Geschlecht der Datenquellen offenzulegen, um die Ergebnisse der Datenanalyse auf eine sexistische Verzerrung überprüfen zu können, dann könnte diese Einbuße für die Privatsphäre der Datenquellen mit ihrer Einwilligung gerechtfertigt sein. Wenn es aber darum geht, die Namen der Datenquellen zu offenbaren oder andere personenbezogene Merkmale, ohne das Einverständnis der Datenquellen einzuholen, dann ist das ethisch problematisch und durch keinen Gewinn an Transparenz für die Datennutzenden aufzuwiegen. Hierbei ist zu beachten, dass es sich bei den risikoexponierten Datenquellen und Datennutzenden um die gleiche Personengruppe handeln kann, weil die Datenquellen selbst an der transparenten Darstellung der möglichen sexistischen Verzerrung interessiert sind. Sobald es sich bei Datenquellen und Datennutzenden aber nicht mehr um die gleiche Gruppe von Personen handelt, ist ein solcher trade-off ethisch nicht mehr zulässig. Denn ein Nachteil für eine Person kann nicht durch einen Vorteil für eine andere Person aufgewogen werden.[1428]

Weiterhin könnte die Beschränkung auf die Risiken der Verletzung der Privatsphäre und der fehlenden Transparenz kritisiert werden. Jobin et al. bemerken passend dazu: „Sustainability, dignity and solidarity are significantly underrepresented compared to other ethical dimensions, which suggests that these issues might be currently flying under the radar of the mainstream AI ethics debate."[1429] Ich argumentiere jedoch, dass die ethischen Dimensionen der Würde und der Solidarität in meinem ethischen Verfahrensmaßstab wiederzufinden sind. Denn es wird die Würde der Datenquellen bewahrt, indem ihnen die Kontrolle über ihre eigenen Daten gewährt wird, und es wird die Würde der Datennutzenden gewahrt, indem sie alle Informationen zur Verfügung gestellt bekommen, die sie brauchen, um algorithmische Ergebnisse einordnen und verstehen zu können. Würde kann ein anderer Ausdruck für die Forderung sein, dass niemand als bloßes Mittel gebraucht werden sollte und ist demnach eine wichtige ethische Dimension in meiner Arbeit. Solidarität

1428 Hansson 2005:10
1429 Jobin et al. 2019:396

könnte hingegen eine alternative Begründung dafür sein, warum insbesondere die Akteure mit der geringsten Kontrolle geschützt und ihre Selbstständigkeit gefördert werden sollte. Ich habe mich hingegen dafür entschieden, mit den Begriffen der Pflicht und der Verantwortung zu argumentieren, da sie eine stärkere Forderung zu stellen scheinen. Der Anspruch an ein solidarisches Miteinander der Akteure lässt sich formal auch mit Kants erster Formulierung des Kategorischen Imperativs begründen, denn das solidarische Miteinander wäre eine verallgemeinerbare Maxime. Mein gesamter Ansatz könnte somit auch mit der Dimension der sozialen Nachhaltigkeit beschrieben werden, da hier ethische Ansprüche gestellt werden, welche das soziale Miteinander im Kontext von Datenökosystem nachhaltig verbessern sollen.

Nur die ethische Dimension der ökologischen Nachhaltigkeit lässt sich in meiner Analyse nicht finden, da dies den Blick auf die Voraussetzungen der Technologie erfordert hätte.[1430] Der Aspekt sei an dieser Stelle noch zur Sprache gebracht: Dadurch, dass es sich bei Datenökosystemen nicht um theoretische Gebilde handelt, sondern die Datenspeicherung und -verarbeitung immer auch Ressourcen verbraucht, lässt sich hier ein weiteres Risiko erkennen, das unabhängig ist von den Beziehungen zwischen den Akteuren. Einerseits gibt es umweltethische Bedenken hinsichtlich der Ressourcen, die bei der Herstellung von Technologien verbraucht werden, insbesondere weil dabei giftige Abfälle entstehen können. Andererseits gibt es umweltethische Bedenken hinsichtlich der Ressourcen, welche die Technologien während ihrer Benutzung verbrauchen – zum Beispiel ist der Stromverbrauch besonders hoch.[1431] Insbesondere der Betrieb von KI-Systemen erfordert enorme Rechenressourcen mit hohem Energieverbrauch und verursacht dadurch zusätzliche Treibhausgasemissionen.[1432] Das Training eines einzigen Deep-Learning-Modells erzeugt ungefähr so viel Kohlendioxid-Emissionen wie fünf Autos während ihrer gesamten Lebensdauer.[1433] Zudem brauchen KI-Systeme enorme Mengen an Wasser zur Kühlung ihrer Prozessoren.[1434]

1430 Bolte, van Wynsberghe 2024:1735f.
1431 Friedman, Kahn 2008:1257
1432 Strubell et al. 2019:3645; Jobin et al. 2019:396; Dodge et al. 2022:1877f.
1433 Van Wynsberghe 2021:213f.; Strubell et al. 2019:3645
1434 Crawford 2024:693

Mein Fokus auf das Datenökosystem als ein soziotechnisches System hat es mir ermöglicht, die Rollen der Akteure zu beleuchten und zu analysieren, wie ein Datenökosystem für die direkt beteiligten Akteure ethisch gestaltet werden kann. Der Blick auf die Nachhaltigkeit des Datenökosystems erfordert aber das Erfassen einer weiteren Dimension – nämlich die technischen Voraussetzungen für das Funktionieren des Datenökosystems. Laut Bolte und van Wynsberghe ist dieser systematische Blick auf die gesellschaftlichen Strukturen, in welche jedes Datenökosystem eingebettet ist, sehr wichtig.[1435] Ich stimme zu, dass die Umstände, in denen Datenökosysteme entstehen, nicht zu vernachlässigen sind. Deshalb ist es möglich, in Schritt drei meines ethischen Verfahrensmaßstabs auch die ökologische Nachhaltigkeit der technologischen Entwicklung als eines der Risiken zu identifizieren. Denn wenn die Menge an verbrauchten Ressourcen aus den Augen verloren wird, entsteht ein langfristiges Risiko bezüglich Ressourcenmangel. In Schritt vier des ethischen Verfahrensmaßstabs können zur Risikominimierung dann Maßnahmen wie Energiesparsamkeit durch Datenminimierung oder neue Kühlmechanismen ohne Wasserverbrauch erarbeitet werden. Somit bietet mein ethischer Verfahrensmaßstab auch den Rahmen für den Schutz und die Förderung der ökologischen Nachhaltigkeit, auch wenn die Entwicklung eines Datenökosystems schon an sich als Risiko für die Umwelt und damit für alle Menschen verstanden werden kann. Insbesondere durch Schritt zwei und drei meines ethischen Verfahrensmaßstabs liegt aber der Fokus auf den zwischenmenschlichen Beziehungen und der Risikoauferlegung innerhalb des Datenökosystems.

Des Weiteren lässt sich bei jedem Ansatz, der helfen soll, technologische Entwicklungen an ethischen Maßstäben auszurichten, die Abhängigkeit von der Einstellung des Unternehmens kritisieren. Denn es liegt im Ermessen des jeweiligen Unternehmens, ob es sich meines ethischen Verfahrensmaßstabs bedient, sich ein ethisches Fundament aufbaut oder schon eines erarbeitet hat und ob es Mechanismen zur Durchsetzung des Verfahrensmaßstabs etabliert.[1436] Bei der

1435 Bolte, van Wynsberghe 2024

1436 Hagendorff 2020:99, 100, 113; Schmitz 2020:369; Munn 2023:871f.; Schmitz 2020:370; Lauer 2021:24

Orientierung an meinem ethischen Verfahrensmaßstab kann, anders als beim Embedded Ethics Ansatz, aber zumindest beurteilt werden, ob das Ethikteam gute Arbeit geleistet hat, indem die vier Schritte anhand des Abschlussberichts nachvollzogen werden. Zudem kann sich das Ethikteam gegen widerstrebende Stimmen im Unternehmen oder gegen widersprüchliche industrielle Interessen durchsetzen, indem es die Relevanz der ethischen Analyse für die Qualität des Produktes betont.[1437]

10.5 Eine Frage der Motivation

Am Ende ist es jedoch eine Frage der Motivation auf Seiten des Unternehmens, ob der hier entwickelte ethische Verfahrensmaßstab angewendet wird. Es spricht viel dafür, sich seiner moralischen Verantwortung anzunehmen und als Unternehmen aktiv die ethische Gestaltung seiner Produkte anzugehen, indem man ein Ethikteam gründet, welches den ethischen Verfahrensmaßstab durchführt. Denn das betreffende Unternehmen übernähme nicht nur negative Verantwortung und würde dementsprechend die Schuld für negative Folgen tragen, sondern es übernähme auch positive Verantwortung und würde deshalb auch für positive Folgen gelobt oder belohnt.[1438] Außerdem ist es eine Tugend, Verantwortung für die eigenen Handlungen und ihre Folgen zu übernehmen[1439] und es kann als Ausdruck von Freiheit gelten, weil die Übernahme von Verantwortung einen freien Willen voraussetzt.[1440] Sie bietet zudem die Möglichkeit, anderen zu helfen, wertvolle moralische Beziehungen zu knüpfen und die eigenen moralischen Fähigkeiten zu entwickeln.[1441] Wenn man auf diese Weise vorausschauende Verantwortung (answerability) übernimmt, wird man laut Kiener als rationaler moralischer Akteur angesprochen, der eine echte moralische Auseinandersetzung wert ist.[1442]

1437 McLennan et al. 2020:490; McLennan et al. 2022:8; Bleher, Braun 2023:7
1438 Nyholm 2024:195
1439 Betzler, Scherrer 2016:13; Wolf 2001:13
1440 Sidgwick 2011 [1874]:50
1441 Kiener 2022:588
1442 Ebd.:591

Allerdings kann es auch Gründe geben, die scheinbar gegen die Orientierung technischer Entwicklungen an ethischen Maßstäben sprechen, wie zum Beispiel die Bequemlichkeit der Datenquellen und Datennutzenden. So stellt sich in Bezug auf die Umsetzung des ethischen Verfahrensmaßstabs die Frage, ob die den Datenquellen und Datennutzenden zugeschriebene Kontrolle in der Praxis von den Betroffenen überhaupt gewürdigt wird. Kant würde dazu sagen: „Faulheit und Feigheit sind die Ursachen, warum ein so großer Teil der Menschen (...) gerne zeitlebens unmündig bleiben; und warum es anderen so leicht wird, sich zu deren Vormündern aufzuwerfen. Es ist so bequem, unmündig zu sein."[1443] Kant meint, den meisten Menschen fehlt der Mut sich ihres eigenen Verstandes zu bedienen. Dadurch seien andere in der Lage, die Kontrolle zu übernehmen und Entscheidungen für diese Menschen zu treffen, ohne sie miteinzubeziehen.[1444] Damit ließe sich argumentieren, dass die meisten Menschen die ihnen in dieser Arbeit zugeschriebene Kontrolle gar nicht wollen. Im Gegenteil: Wenn mehr Kontrolle auf Seiten der Datenquellen und Datennutzenden die digitale Dienstleistung komplizierter macht, lehnen Datenquellen und Datennutzende die Kontrolle eventuell sogar aus Bequemlichkeit ab.

Aus utilitaristischer Perspektive könnte hinzugefügt werden, dass es den Menschen, solange sie sich der Risikoauferlegung nicht bewusst sind und der riskierte Schaden auch nicht eintritt, womöglich sogar egal ist, ob sie als bloßes Mittel gebraucht werden. Wer keinen Schaden spürt, hat keinen Nachteil und will möglichst wenig Arbeit mit digitalen Diensten haben, denn dieser Aufwand wäre ja wiederum ein Nachteil. Die tadelnswerte Ignoranz gegenüber den Risiken könnte allerdings der Unwissenheit geschuldet sein. Denn nur weil der Schaden nicht gespürt wird, entsteht er ja trotzdem. Durch jegliche Risikoauferlegung entsteht ein Schaden an der Autonomie der Risikoexponierten, wie in Kapitel 5 gezeigt. Zudem können die potenziellen Folgen der Risiken nicht gewollt werden: Die Verletzung der Privatsphäre kann zu Erpressung, staatlicher Kontrolle oder kommerzieller Ausbeutung führen[1445] und im Endeffekt sogar zu einem Fehlen von autonomer Entfaltung, intimen

1443 Kant 1784:481
1444 Ebd.
1445 Hare 1993:15

Beziehungen der Liebe und der Freundschaft und damit einem fehlenden Wohlbefinden.[1446] Das Fehlen von Transparenz bedeutet fehlendes Verständnis von der Entstehung eines Ergebnisses und das führt zu der Unfähigkeit reflektiert Stellung dazu zu nehmen und das Ergebnis anzufechten.[1447] So können Datennutzende keine fundierte Entscheidung darüber treffen, wie sie mit den betreffenden Ergebnissen umgehen[1448], und können ihre eigenen Interessen nicht vertreten[1449].

Es ist für Datenquellen ein vergleichsweise geringer Aufwand sich über ihre Privatsphäre Gedanken zu machen und zu kontrollieren, mit wem sie welche Daten teilen möchten, im Kontrast zu dem garantierten Schaden durch die Risikoauferlegung und den potenziellen negativen Folgen bei der Verletzung der Privatsphäre. Gleichermaßen ist es für Datennutzende ein vergleichsweise geringer Aufwand die algorithmischen Ergebnisse in ihrem Kontext zu verstehen, anstatt sich einfach auf ihre Fehlerlosigkeit zu verlassen, im Kontrast zu dem garantierten Schaden durch die Risikoauferlegung und den potenziellen negativen Folgen beim Fehlen der Transparenz. Somit überwiegt sogar aus utilitaristischer Perspektive der (potenzielle) Schaden, solange man sich dessen bewusst ist. Wenn die Menschen über die Schäden aufgeklärt werden und ihnen bewusst wird, dass sie schon allein durch die Risikoauferlegung auf Grund der Beschränkung ihrer Autonomie einen Schaden erleiden, wollen sie die Wahl haben. Die in dieser Arbeit entwickelte deontologische Risikoethik und mein ethischer Verfahrensmaßstab verschaffen den Datenquellen und Datennutzenden die Wahl und verlangen ihre Aufklärung – es ist aber niemand gezwungen dieses Angebot auch anzunehmen.

Trotzdem könnte sich die Bequemlichkeit der Datenquellen und Datennutzenden auf die Umsetzung des in dieser Arbeit entwickelten ethischen Verfahrensmaßstabs auswirken. Angenommen ein digitaler Dienst wird komplizierter durch die Umsetzung des ethischen Verfahrensmaßstabs und sowohl Datenquellen als auch Datennutzende wollen sie deshalb aus Bequemlichkeit nicht mehr benutzen, dann ist es für die Anbieter der digitalen Dienste nicht

1446 Rössler 2003:30; Moore 2010: 21; 56, 87; Fried 1968:216
1447 Keymolen, Taylor 2023:485f.
1448 Munch et al. 2024:9
1449 Vredenburgh 2022:212

attraktiv den ethischen Verfahrensmaßstab umzusetzen. Auf diese Weise könnte die Bequemlichkeit der Datenquellen und Datennutzenden, ihre eigene Kontrolle in Anspruch zu nehmen, dazu führen, dass Unternehmen, die wettbewerbsfähig bleiben wollen, davon absehen den ethischen Verfahrensmaßstab umzusetzen.

Hierbei ist die Annahme, dass man sich systematisch schlechter stellt, wenn man ethisch gestaltet. Ich möchte allerdings in Frage stellen, ob die Umsetzung meines ethischen Verfahrensmaßstabes die Nutzung digitaler Dienste überhaupt komplizierter macht. Womöglich führt die Minimierung der identifizierten Risiken in Schritt 4 zu Datenschutzmaßnahmen, Datensicherheitsmaßnahmen oder zu ausführlicheren algorithmischen Ergebnissen, die den Kontext deutlich machen. Aber jede dieser Maßnahmen führt nur zu einem Angebot. Wie derzeit schon bei Browsercookies üblich, gibt man den Datenquellen und Datennutzenden ausschließlich die Möglichkeit zuzustimmen oder abzulehnen. Wer zu bequem ist, sich damit auseinanderzusetzen, was das Setzen von Cookies bedeutet und ob man das möchte, der kann einfach auf „Alle akzeptieren" klicken und wird in der Nutzung des digitalen Dienstes nicht weiter gestört. Wer sich aber dafür interessiert, welche Informationen über die eigenen Internet-Aktivitäten gesammelt werden und somit nachvollziehen können will, wie zum Beispiel personalisierte Werbung zustande kommt, der kann sich die Zeit nehmen die Informationen zu lesen und durch das Einschränken der Cookies die eigenen Interessen vertreten oder es sich zur Gewohnheit machen, sie, so weit möglich, abzulehnen, und es sich damit einfacher machen.

Für die Anbieter des digitalen Dienstes – für die Datengenerierenden und Datenverarbeitenden des Datenökosystems – macht mein ethischer Verfahrensmaßstab das Gestalten der Dienste womöglich zunächst aufwendiger, wenn zuvor alle Dienste ohne Reflexion über potenzielle gesellschaftliche Auswirkungen gestaltet wurden. Trotzdem wird die Gestaltung dadurch nicht komplizierter. Stattdessen wird durch meinen ethischen Verfahrensmaßstab klarer, welche Schritte absolviert werden müssen, um ein Datenökosystem oder seine einzelnen technischen Elemente ethisch zu gestalten. Beim ersten Mal bedeutet die Umsetzung des ethischen Verfahrensmaßstabes natürlich einen Mehraufwand für die Anbieter durch die Einstellung der Ethikerin oder des Ethikers und auf Grund der Zeitverzögerung durch den Reflexionsprozess. Allerdings wird durch diesen Mehr-

aufwand das Produkt besser und kundenfreundlicher. Denn indem man ethische Überlegungen im Vorhinein anstellt und Risiken antizipiert, passieren weniger vermeidbare Fehler. So wird das Produkt sicherer und die Anbieter müssen im Idealfall der Öffentlichkeit keinen Datenverlust oder fehlerhafte Ergebnisse erklären. Insofern lohnt sich der Reflexionsprozess am Ende und macht letztlich das Produkt sicherer und kostengünstiger.

Dabei ist es essenziell, dass Datengenerierende und Datenverarbeitende sich bewusst sind, warum es sich lohnt ihre Produkte ethisch zu gestalten, damit sie einen Anreiz haben, meinen ethischen Verfahrensmaßstab umzusetzen. Ich meine damit nicht, dass sich ethische Gestaltung finanziell lohnen muss, um attraktiv zu sein – obwohl man sich als Anbieter auch profilieren kann mit der besonders ethischen Gestaltung seiner Produkte und daraus sicher auch Gewinn ziehen kann. Sondern ich meine, dass es sich für ein angenehmes gesellschaftliches Miteinander lohnt, auch jeden technischen Dienst ethisch zu gestalten. Angenommen ein Unternehmen möchte gute technische Produkte hervorbringen – und davon gehe ich aus – ist die Qualität des Produktes nicht nur davon abhängig, wieviel Geld sie dem Unternehmen einbringt, sondern auch welchen Beitrag es zum gesellschaftlichen Miteinander leistet.

Wen es nicht reizt, sich darüber Gedanken zu machen, wie andere wohl gerne behandelt werden würden, der sollte sich bewusst machen, dass er/sie selbst auch, womöglich in anderen Kontexten, in der Rolle der Datenquelle oder des Datennutzenden sein wird – wie würde man dann selbst gerne behandelt werden? Mein ethischer Verfahrensmaßstab, auf Grundlage der deontologischen Risikoethik, ermöglicht einen gesellschaftlichen Umgang auf Augenhöhe durch das Ausgleichen der Kontrolle der Akteure im Datenökosystemen und durch das Lenken der Aufmerksamkeit auf die autonomiebeschränkende Risikoauferlegung. Auch die potenziellen negativen Folgen beim Risikoeintritt werden antizipiert, sodass sie mit Hilfe der administrativen Anforderungen vermieden werden können.

Um einen zusätzlichen Anreiz für Unternehmen zu schaffen, seine Produkte ethisch zu gestalten, könnte ein diesbezügliches Zertifikat erstellt und gegebenenfalls zugewiesen werden. Dieses Zertifikat kann dann zum Nachweis der moralischen Integrität in der Öffentlichkeit verwendet werden. Hierfür müsste es aber eine zentrale

Institution geben, welche das Vorgehen des Ethikteams mit Hilfe von unabhängigem Personal überprüft und die Berechtigung hat, bei der Erfüllung bestimmter Bedingungen ein Zertifikat für ethische Gestaltung auszustellen. Zur Kontrolle der Einhaltung der DSGVO gibt es zum Beispiel Datenschutzbehörden – solche Aufsichtsbehörden könnte es auch für die Kontrolle der ethischen Gestaltung geben. Es würde jedoch einen erheblichen Aufwand bedeuten, den Entwicklungsvorgang jedes Produktes aller Unternehmen zu überprüfen. So bleibt die ethische Gestaltung erst einmal etwas, das um seiner selbst willen von den Unternehmen angestrebt werden sollte. So kann auch ohne äußeres Kontrollgremium auf Grundlage innerer Selbstregulierung der Unternehmen die Entwicklung von Technologien und ganzen Datenökosystemen an ethischen Maßstäben orientiert werden.

Abschließend möchte ich argumentieren, dass die Analyse in dieser Arbeit und der daraus resultierende Vorschlag für eine ethische Gestaltung von Datenökosystemen unabhängig sind von den Vorlieben der Akteure. Denn es spielt keine Rolle, ob die betroffenen Datenquellen und Datennutzenden die Kontrolle zu schätzen wissen, die ihnen in dieser Arbeit zugeschrieben wird. Sie haben aus ethischer Sicht sowieso einen Anspruch darauf. Nur weil die Akteure zu bequem sein mögen, um die ihnen zustehende Kontrolle auszuüben (deskriptiv), heißt das nicht, dass sie ihre Kontrolle nicht ausüben sollten (normativ) – das wäre ein naturalistischer Fehlschluss.[1450] In gleicher Weise besteht auch der normative Anspruch an die Datengenerierenden und Datenverarbeitenden, Datenökosysteme mit Hilfe meines ethischen Verfahrensmaßstabs ethisch zu gestalten, unabhängig von Sorgen über die Wettbewerbsfähigkeit, die dagegen zu sprechen scheinen.

Diese Arbeit hat natürlich trotzdem nur dann einen Effekt auf unser gesellschaftliches Zusammenleben, wenn sie auch als richtig anerkannt wird und wenn gewissen Widerstände – Bequemlichkeit sowie zusätzliche Zeit und Kosten – zum Zwecke der Umsetzung der hier angestellten Überlegungen überwunden werden.

1450 Hume 2000[1739]:Buch III, Teil I, Kapitel I

11 Fazit für die ethische Gestaltung im Umgang mit digitalen Daten

In dieser Arbeit wurde gezeigt, wie der Umgang mit digitalen Daten in Datenökosystemen ethisch gestaltet werden kann. Begonnen wurde mit der Erkenntnis, dass Datenökosysteme als ein System von Akteuren, zwischen denen mit Hilfe von technologischen Mitteln Daten fließen, verstanden werden kann. Allein auf Grund der Struktur, den Beziehungen zwischen den Akteuren, werden ihnen unterschiedlich viel Wissen und Kontrolle über den Datenfluss im Datenökosystem ermöglicht. *Datenquellen* stellen ihre Daten zur Verfügung, *Datengenerierende* sammeln und speichern die Daten, *Datenverarbeitende* kombinieren und analysieren die Daten, um sie dann aufbereitet als Ergebnisse zur Verfügung stellen zu können, und *Datennutzende* arbeiten mit den aufbereiteten Daten, bzw. sie sind auf diese Daten angewiesen. Datengenerierende und Datenverarbeitende sind dabei die Akteure mit mehr Wissen und Kontrolle, da sie wissen und bestimmen können, was mit den Daten passiert. Dadurch wurde klar, dass der Umgang mit digitalen Daten im Datenökosystem eine Risikoauferlegung für Datenquellen und Datennutzende beinhaltet. Denn Datenquellen können nicht kontrollieren, was mit ihren Daten geschieht, sodass ihnen bei der Datensammlung das Risiko der Verletzung ihrer Privatsphäre und somit ein Verlust von Autonomie auferlegt wird. Datennutzende können hingegen nicht nachvollziehen, wie die Ergebnisse der Datenverarbeitung zustande gekommen sind, sodass ihnen bei der intransparenten Bereitstellung der Ergebnisse das Risiko auferlegt wird, auf dieser Grundlage zu falschen Überzeugungen zu gelangen und somit nicht für ihre Interessen einstehen zu können.

Da diese Risikoauferlegung, sobald sie entgegen der Wünsche der Risikoexponierten geschieht, deren Autonomie beschränken kann, entsteht ein ethisches Problem. Um den Umgang mit Daten im Datenökosystem trotzdem ethisch gestalten zu können, wurde in dieser Arbeit erörtert, wann es zulässig ist, anderen ein Risiko auf-

zuerlegen – nämlich nur dann, wenn die Risikoauferlegung Teil eines gerechten sozialen Systems der Risikobereitschaft ist, von dem alle profitieren und in dem niemand als bloßes Mittel gebraucht wird. Der Kategorische Imperativ von Kant verbietet aber nicht nur den Gebrauch anderer als bloßes Mittel, sondern auch bestimmte Handlungsfolgen, die durch den Umgang mit Daten in einem Datenökosystem riskiert werden: die Verletzung der Privatsphäre und das Fehlen von Transparenz.

Es ließ sich in dieser Arbeit argumentativ nachweisen, dass die Privatsphäre der Datenquellen durch den Zugriff auf ihre Daten ohne ihre Wahl-Kontrolle verletzt werden kann. Diese Verletzung der Privatsphäre würde bedeuten, dass die Datenquellen der Gefahr von Erpressung, staatlicher Kontrolle oder kommerzieller Ausbeutung sowie Autonomieverlust ausgesetzt wären. Das wäre ein unerwünschtes Ereignis, da dies zu ungewollten Übergriffen oder Diskriminierung führen kann und weil eine intakte Privatsphäre eine unbedingte Voraussetzung für ein autonomes Leben ist. Da dieses unerwünschte Ereignis der Privatsphärenverletzung nur mit einer gewissen Wahrscheinlichkeit eintritt, kann sie als Risiko bezeichnet werden.

Außerdem konnte argumentativ nachgewiesen werden, dass Datennutzenden auf Grund der epistemischen Asymmetrie in Datenökosystemen Transparenz verwehrt werden kann, indem die Information, wie die Ergebnisse der Datenverarbeitung zustande gekommen sind, verwehrt wird. Die Unmöglichkeit, die Ergebnisse zu überprüfen, ist aus einer epistemischen Perspektive problematisch, weil verzerrte Ergebnisse zu falschen Überzeugungen führen können. Dies sollte aus ethischer Sicht vermieden werden, da es für das menschliche Wohlergehen wichtig ist, sich von richtigen und nicht von falschen Überzeugungen leiten zu lassen und das Wissen darüber notwendig ist, um seine eigenen Interessen vertreten zu können. Da es bei fehlender Transparenz jedoch möglich ist, dass Datennutzende vorsätzlich falsch informiert oder durch verzerrte Ergebnisse unabsichtlich unfairen Entscheidungen ausgesetzt werden, handelt es sich hier ebenfalls um unerwünschte Ereignisse, die mit einer gewissen Wahrscheinlichkeit eintreten können, und damit um ein Risiko.

Das heißt, die Privatsphäre der Datenquellen und die Transparenz für die Datennutzenden können durch die Datenverarbeitung im

Datenökosystem gefährdet werden – das sollten sie aber nicht. Es kann nicht gewollt werden, dass Datenquellen und Datennutzenden diese Risiken auferlegt werden, deshalb entsteht eine moralische Pflicht für Datengenerierende und Datenverarbeitende, dies zu vermeiden. Da aber eine Auferlegung von Risiken nicht immer vermeidbar ist, kann mit Hilfe der in dieser Arbeit entwickelten deontologischen Risikoethik erörtert werden, unter welchen Bedingungen sie trotzdem ethisch zulässig sein kann. So stellt sich heraus, dass durch die Einhaltung einiger administrativer Anforderungen die Privatsphäre der Datenquellen und die Transparenz für die Datennutzenden im Datenökosystem geschützt und gefördert werden können:

Administrative Anforderungen für die Umsetzung von Privatsphäre:

- *Einverständnis* der Individuen wird eingeholt
- Gewährleistung von *Freiwilligkeit* der Einwilligung
- Individuen werden über alle Verwendungszwecke *informiert*
- Gewährleistung von *Zweckentsprechung* der Informationen

Administrative Anforderungen für die Umsetzung von Transparenz:

- *Entstehung* des Ergebnisses verstehen: Welcher Input führt zu welchem Output und welche Trainingsdaten wurden verwendet?
- Algorithmische Ergebnisse im zum Verständnis notwendigen *Kontext* darstellen.

Da die Auferlegung eines Risikos nur dann zulässig ist, wenn niemand als bloßes Mittel gebraucht wird und die administrativen Anforderungen dafür sorgen, dass die Selbstbestimmung der Akteure ermöglicht wird und sie somit nicht als bloßes Mittel gebraucht werden, ist die Erfüllung der administrativen Anforderungen eine notwendige Bedingung für die Zulässigkeit der Risikoauferlegung.

Als nächstes wurde gezeigt, dass aus einer Position der erhöhten Kontrolle über die Datenverarbeitung innerhalb eines Datenökosystems Verantwortung erwächst für die Minimierung der Risiken und das Erzeugen von positiven Ergebnissen. Denn es ließ sich argumentativ nachweisen, dass Datengenerierende und Datenverarbeitende auf Grund ihrer erhöhten Kontrolle innerhalb des Datenökosystems Verantwortung tragen für den Schutz und die Förderung der Privatsphäre der Datenquellen und der Transparenz für die Datennut-

zenden. Somit ist offensichtlich, wer für die ethische Gestaltung von Datenökosystemen zuständig ist. Wie diese Gestaltung aussehen soll, wurde anhand der aus der Arbeit abgeleiteten Kriterien für die ethische Gestaltung erwogen. Nur die Ethischen Leitlinien und die ethische Risikoanalyse von Hansson erfüllen alle Kriterien für die ethische Gestaltung einer Technologie:

i) Die Umsetzungsmethode sollte konkrete praktische Anleitung bieten.
ii) Die Umsetzungsmethode sollte den Einfluss der Technologie auf alle Betroffenen berücksichtigen.
iii) Im Rahmen der Umsetzungsmethode sollte die Verantwortung für die Umsetzung der ethischen Überlegungen bestimmten menschlichen Akteuren zugeordnet werden.

Erst wenn diese vier Kriterien alle erfüllt sind, kann in Betracht gezogen werden, dass der jeweilige Ansatz zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben auch für die ethische Gestaltung eines Datenökosystems geeignet ist. Allerdings erfüllen die beiden Ansätze, die sich zur Gestaltung einer Technologie eignen würden, nicht alle Kriterien für die ethische Gestaltung eines Datenökosystems:

I. Die Umsetzungsmethode sollte es ermöglichen, die Auferlegung von Risiken zu regulieren.
II. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen zu erfüllen.
III. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen.

Denn die Umsetzungsmethode der Ethischen Leitlinien ermöglicht es nicht, die Risikoauferlegung zu problematisieren und dementsprechend die Auferlegung von Risiken in der Praxis regulieren (I). Die ethische Risikoanalyse ermöglicht es zwar die Risikoauferlegung zu problematisieren, aber sie beantwortet die Frage, unter welchen Umständen das Risiko zulässig ist, nur mit einem Risikotausch, der für alle Beteiligten von Vorteil ist. Da es für eine zulässige Risikoauferlegung nicht reicht, wenn ein Risiko durch einen eigenen Vorteil kompensiert wird, reicht diese Antwort auch nicht, um die Auferlegung

von Risiken regulieren zu können (I). Deshalb habe ich aus den Erkenntnissen der bisherigen Arbeit und auf Grundlage der ethischen Risikoanalyse von Hansson einen eigenen ethischen Verfahrensmaßstab entwickelt, der die Frage beantwortet, wie Datenökosysteme gestaltet werden sollen. Durch die Anwendung der folgenden vier Schritte ist es den Argumenten in dieser Arbeit zufolge möglich, ein Datenökosystem ethisch zu gestalten:

Schritt 1: Positionierung zu Verantwortung, Nutzen und Risiken
Schritt 2: Risikoverteilung bewerten mit Hanssons ethischer Risikoanalyse
Schritt 3: Überprüfung der Zulässigkeit von Risikoauferlegung
Schritt 4: Fehlervorsorge durch Minimierung der Risiken

Aus der ethischen Erkenntnis, dass Risiken nur dann auferlegt werden dürfen, wenn dadurch niemand als bloßes Mittel gebraucht wird, und aus den Konzepten der Privatsphäre und der Transparenz abgeleiteten administrativen Anforderungen wurde eine Umsetzungsmethode entwickelt, die es ermöglicht, bei der Gestaltung eines Datenökosystems und seiner technischen Einzelelemente die ethischen Anforderungen einzubinden.

Die beiden Forschungsfragen aus der Einleitung wurden in dieser Arbeit beantwortet:

1. Wie lassen sich Datenökosysteme ethisch bewerten?
2. Wie sollten Datenökosysteme dementsprechend gestaltet werden?

Zur Beantwortung der ersten Forschungsfrage wurde argumentiert, dass sich Datenökosysteme am besten mit Hilfe der deontologischen Risikoethik ethisch bewerten lassen. Dies war möglich, weil anhand der Struktur von Datenökosystemen erkennbar wurde, dass die Akteure unterschiedlich viel Wissen und Kontrolle haben und zudem das Sammeln und Verarbeiten von Daten im Datenökosystem als Risikoauferlegung verstanden werden muss. Von den drei dargestellten klassischen Moraltheorien hat es nur die kantianische Deontologie ermöglicht:

a. das asymmetrische Wissen sowie die ungleiche Kontrolle der Akteure im Datenökosystem zu betrachten und
b. das Auferlegen der Risiken zu bewerten.

Im Datenökosystem ist es also dann zulässig, Datenquellen und Datennutzenden Risiken aufzuerlegen, wenn sie dabei nicht als bloßes Mittel gebraucht werden. Das heißt, solange die individuelle Kontrolle so groß wie möglich ist, weil die Risikoexponierten ohne Täuschung über das Risiko informiert werden und ohne Zwang die Wahl haben, ob sie das Risiko tragen wollen, lassen sich Datenökosysteme ethisch als zulässig einstufen.

Die zweite Forschungsfrage wurde mit der Entwicklung des ethischen Verfahrensmaßstabs beantwortet. Datenökosysteme sollten entsprechend meines ethischen Verfahrensmaßstabs gestaltet werden, da dieser Vorschlag zur Umsetzung digitaler Ethik in die Praxis gut geeignet ist, um ethische Probleme im Zusammenhang mit der Datenverarbeitung in Datenökosystemen zu adressieren. Denn durch meinen ethischen Verfahrensmaßstab können einerseits die Risiken des Datenökosystems minimiert werden, andererseits kann dadurch die aktive Gestaltung von Datenökosystemen an der deontologischen Risikoethik orientiert werden, um die Ungleichheit des Wissens und der Kontrolle über Risiken möglichst auszugleichen.

Meine Arbeit leistet in dreierlei Hinsicht einen neuen Beitrag zur Forschung im Bereich der digitalen Ethik:

- mit der Entwicklung der deontologischen Risikoethik,
- mit dem Erarbeiten der administrativen Anforderungen und
- mit dem daraus abgeleiteten ethischen Verfahrensmaßstab.

Mit Kants dritter Formel des Kategorischen Imperativs habe ich eine Antwort gefunden auf die Frage, wann es zulässig ist, anderen ein Risiko aufzuerlegen. Das Prinzip von Hansson war dabei eine wichtige Grundlage. Ich baue darauf auf, dass jede Risikoauferlegung nur zulässig ist als Teil eines gerechten sozialen Systems der Risikobereitschaft, das auch den Risikoexponierten zum Vorteil gereicht.[1451] Dieses System soll laut Hansson Regeln enthalten, die bestimmte Arten von risikobehafteten Aktivitäten zulassen – er stellt jedoch nur die Regel auf, dass die Risikoauferlegung für alle Beteiligten von Vorteil muss. Da ich aber der Ansicht bin, dass es für eine zulässige Risikoauferlegung nicht reicht, wenn ein Risiko durch einen eigenen Vorteil kompensiert wird, habe ich die dritte Formel des Kategori-

1451 Hansson 2003:305

schen Imperativs von Kant als eine weitere Regel vorgeschlagen. Dadurch ist es nun möglich, eine klare Antwort auf die Frage zu geben, unter welchen Bedingungen es zulässig ist, andere einem Risiko auszusetzen. Hierdurch lassen sich auch konkrete administrative Anforderungen zum Schutz vor bestimmten Risiken ableiten.

Die administrativen Anforderungen zur Umsetzung von Privatsphäre und Transparenz wurden in dieser Arbeit aus den ausführlichen Definitionen dieser Konzepte und dem normativen Anspruch an ihren Schutz abgeleitet. Durch die Begriffsdefinitionen wurde klar, unter welchen Umständen Privatsphäre intakt bzw. verletzt ist und wie Transparenz überhaupt ermöglicht werden kann bei der Bereitstellung von algorithmischen Ergebnissen. Dank der Begründung einer Pflicht zum Schutz der Privatsphäre und der Bereitstellung von transparenten Ergebnissen konnte geschlussfolgert werden, dass sowohl der Schutz der Privatsphäre als auch die Bereitstellung von transparenten Ergebnissen in der Praxis gewährleistet werden sollten. Die administrativen Anforderungen liefern also eine Antwort auf die Frage, unter welchen Bedingungen die Privatsphäre intakt und die Transparenz gesichert ist. Somit sind die Anforderungen an eine ethische Gestaltung gut begründet.

Unter anderem dadurch wurde auch der Umsetzungsvorschlag für die Praxis in dieser Arbeit ausführlich theoretisch hergeleitet. Denn die Kriterien für die ethische Gestaltung eines Datenökosystems wurden aus der deontologischen Risikoethik und den administrativen Anforderungen abgeleitet. Zuvor sollten aber noch die allgemeineren Kriterien für die ethische Gestaltung von Technologien erfüllt sein. Auch diese stammen aus gut begründeten theoretischen Überlegungen zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben. Denn diese ist nur möglich, wenn konkrete praktische Anleitung gegeben ist, es spezifische und gut begründete ethische Anforderungen an die Gestaltung gibt, der Vorteil für alle von der Technologie betroffenen Personen das Ziel ist und die Verantwortung für die ethische Gestaltung des Datenökosystems einem bestimmten menschlichen Akteur zugeordnet ist. Dadurch konnte ein ethischer Verfahrensmaßstab auf Grundlage der ethischen Risikoanalyse von Hansson aufgestellt werden, der konkrete praktische Anleitung bietet, dabei den Einfluss der Technologie auf alle Betroffenen berücksichtigt und für dessen Anwendung das zu diesem Zweck aufzustellende Ethikteam verantwortlich ist. Dank diesem

ethischen Verfahrensmaßstabs ist es auch möglich, die Auferlegung von Risiken zu regulieren und die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen und zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen. Somit wurde in dieser Arbeit nicht nur ein theoretischer Beitrag geleistet, sondern es wurde mit dem ethischen Verfahrensmaßstab auch eine Möglichkeit zur praktischen Umsetzung dieser theoretischen Überlegungen geschaffen.

In dieser Arbeit habe ich mich der besten Eigenschaften der von Bolte und van Wynsberghe sogenannten zweiten und dritten Welle der KI-Ethik bedient. Einerseits beschäftige ich mich, wie die zweite Welle, mit den konkreten Risiken, die aktuell durch den Umgang mit Daten entstehen. Andererseits nutze ich, wie die dritte Welle, zur Analyse dieser Risiken einen strukturellen Ansatz. Denn es ging in dieser Arbeit nicht um einzelne Technologien, sondern um die Struktur von Datenökosystemen, in denen sie angewendet werden. Dabei verfolge ich, ebenfalls wie die dritte Welle, einen systemisch orientierten Ansatz, der ethische Fragestellungen auf struktureller Ebene sichtbar macht. So wurde deutlich, dass der Umgang mit Daten im Datenökosystem als Risikoauferlegung verstanden werden kann und sich dementsprechend die Frage stellt, wann eine solche Risikoauferlegung ethisch zulässig ist. Um diese ethischen Überlegungen zu realisieren, setze ich, wie die zweite Welle, auf einen durchdachten Gestaltungsprozess, um sozio-technische Strukturen zu verändern. Dabei gehe ich nicht davon aus, dass sich ethische und strukturelle Probleme allein durch Verbesserungen einzelner Anwendungen lösen lassen und distanziere mich somit von dem Techno-Solutionismus der zweiten Welle – vielmehr lassen sich die Probleme durch Kommunikation über die Zwecke der Datenverarbeitung und die Aussagekraft der Ergebnisse lösen. Ebenso vertrete ich auch nicht die Auffassung der dritten Welle, dass erst umfassende strukturelle Veränderungen notwendig sind, bevor Technologie ethischer gestaltet werden kann – es ist schon jetzt möglich, Datenquellen und Datennutzende nicht im Ungewissen über Pläne und Vorgehensweisen zu lassen.

Das OZG und die laufende Digitalisierung der Verwaltungsdienstleistungen in Deutschland bieten in dieser Arbeit hilfreiche Beispiele zur Veranschaulichung der theoretischen Überlegungen. Das bedeu-

tet jedoch nicht, dass die erarbeitete Theorie sowie der ethische Verfahrensmaßstab ausschließlich zur Anwendung im Verwaltungskontext geeignet sind. Der ethische Verfahrensmaßstab kann für die Gestaltung jeglicher Datenökosysteme angewandt werden. Dabei ist jedoch zu beachten, dass die Gruppenprivatsphäre womöglich neue administrative Anforderungen an die Gestaltung verlangt. Denn sobald Datenverarbeitende Gruppenprofile erstellen, um Werbung oder die Bepreisung von Produkten personalisieren zu können, reicht die Berücksichtigung der individuellen Privatsphäre nicht mehr zum Schutz vor Ungleichbehandlung. Das heißt, je nach Kontext des Datenökosystems, sollte die Gruppenprivatsphäre beim Minimieren des Risikos in Schritt 4 des ethischen Verfahrensmaßstabs Beachtung finden.

In unterschiedlichen Kontexten mag es unterschiedliche Probleme bei der Umsetzung von Bedingungen für ein gerechtes soziales System der Risikobereitschaft geben. Zum Beispiel ist im Kontext medizinischer Forschung die von mir geforderte individuelle Einwilligung oft nur schwer umsetzbar. Während im Verwaltungskontext jeder Antrag einzeln bearbeitet werden muss und es somit möglich ist, jede antragstellende Person nach der Einwilligung zur Nutzung ihrer Daten zu fragen, ist es in der medizinischen Forschung wichtig, so viele Daten wie möglich zu sammeln, damit ihre Analyse aussagekräftig ist. Das heißt, die administrativen Anforderungen zum Schutz der Privatsphäre sind im Kontext der Verwaltung umsetzbar, können jedoch in anderen Kontexten Anpassungen erfordern. In jedem Fall gilt das Prinzip der Maximierung der individuellen Kontrolle – seine Umsetzung kann aber in verschiedenen Kontexten unterschiedlich ausgestaltet werden. Das heißt, bei der Anwendung meines ethischen Verfahrensmaßstabs in anderen Kontexten, wie zum Beispiel medizinischer Forschung, können im Rahmen der Kontrollmaximierung durch Informationen und Kompensation Lösungen gefunden werden.

In Bezug auf die Transparenz wurde in dieser Arbeit zwar davon ausgegangen, dass der Einsatz von Black Box KI-Algorithmen (KI-Systeme) für die Datennutzenden ein besonders großes Risiko erzeugt, aber auch wenn die bereitgestellten Ergebnisse für die Datenverarbeitenden nachvollziehbar sind, gelten die gleichen administrativen Anforderungen an die Umsetzung der Transparenz. Insbesondere wenn sich die Ergebnisse auf Personen beziehen, ist ihre

Nachvollziehbarkeit ethisch geboten, damit Einspruch erhoben und die eigenen Interessen vertreten werden können. Aus epistemischer Sicht ist aber auch unabhängig von Personen die Transparenz essentiell für jeglichen Erkenntnisgewinn. Das heißt, die Anforderungen an die Transparenz können auch hilfreich sein für Datenökosysteme, die sich nicht auf Personen beziehen.

Auch die in dieser Arbeit entwickelte deontologische Risikoethik lässt sich über den Verwaltungskontext und Datenökosysteme hinaus auf jegliche Risikoauferlegung anwenden. Immer wenn sich die Frage stellt, unter welchen Bedingungen es zulässig ist, andere einem Risiko auszusetzen, bietet die deontologische Risikoethik eine Antwort. Denn aus meiner Sicht ist ein System der Risikobereitschaft nur dann gerecht und sozial und gereicht allen zum Vorteil, wenn die Regel gilt, dass niemand als bloßes Mittel gebraucht wird. Deshalb kann dieser Ansatz in jeglichem Kontext hilfreich sein, wenn es darum geht, die ethische Zulässigkeit von Risikoauferlegung einzuschätzen.

Diese Arbeit bietet eine Grundlage, um bewusst zu reflektieren, welche Ziele bei der Gestaltung von Datenökosystemen gefördert werden sollten. Sie liefert damit einen normativen Orientierungsrahmen für die ethische Gestaltung von Datenökosystemen – jenseits technischer Einzellösungen und unter Berücksichtigung struktureller Bedingungen. Somit trägt diese Arbeit mit Hilfe von vertieften theoretischen Erörterungen dazu bei, technische Entwicklungen in der Praxis an ethischen Maßstäben auszurichten. Es lässt sich aus den Ergebnissen schließen, dass es unter bestimmten Bedingungen möglich ist, Datenökosysteme ethisch zu gestalten und wie diese Bedingungen hergestellt werden können. Die Umsetzung dieser Bedingungen ermöglicht es zum Beispiel, dass wir online einen neuen Ausweis beantragen oder unseren Wohnsitz ummelden können, und uns die dabei entstehenden Risiken nur unter zulässigen Bedingungen auferlegt und unsere Privatsphäre sowie die Transparenz der Ergebnisse geschützt und gefördert werden.

Literaturverzeichnis

Abomhara, Mohamed und Geir M. Køien (2015): „Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders and Attacks", *Journal of Cyber Security*, Bd. 4, S. 65–88, doi: 10.13052/jcsm2245-1439.414.

Acquisti, Alessandro; Laura Brandimarte und Jeff Hancock (2022): „How privacy's past may shape its future – An account of privacy's evolutionary roots may hold lessons for policies in the digital age", *Science*, Bd. 375(6578), S. 270–272, DOI: 10.1126/science.abj0826.

Adner, Ron (2017): „Ecosystem as Structure: An Actionable Construct for Strategy", *Journal of Management*, Bd. 43(1), S. 39–58, DOI: 10.1177/0149206316678451.

Alston, William P. (2005): „Beyond "Justification" – Dimensions of Epistemic Evaluation", Cornell University Press.

Altintac, Ismail (2022): „Digitalisierung und Industrie 4.0", in: Meier, Klaus-Jürgen und Matthias Pfeffer (Hrsg.): „Produktion und Logistik in der digitalen Transformation – Analyse – Planung – Praxiserfahrungen", Springer Gabler, S. 3–24.

Altman, Irwin (1976): „Privacy – A conceptual analysis", *Environment and Behavior*, Bd. 8(1), S. 7–29, https://doi.org/10.1177/001391657600800102.

Ariss, Steven M. (2009): „Asymmetrical knowledge claims in general practice consultations with frequently attending patients: Limitations and opportunities for patient participation", *Social Science & Medicine*, Bd. 69, S. 908–919, https://doi.org/10.1016/j.socscimed.2009.06.045.

Aristoteles (2006): „Nikomachische Ethik", Hrsg. und übersetzt von Ursula Wolf, Rowohlt Verlag.

Asgarinia, Haleh (2023): „Convergence of the source control and actual access accounts of privacy", *AI and Ethics*, Bd. 4, S. 333–343, https://doi.org/10.1007/s43681-023-00270-z.

Attard, Judie; Fabrizio Orlandi und Sören Auer (2016): „Data value networks: enabling a new data ecosystem", in: 2016 IEEE/WIC/ACM international conference on web intelligence (WI), IEEE, S. 453–456, DOI 10.1109/WI.2016.0073.

Auernhammer, Jan (2020): „Human-centeres AI: The role of Human-centeres Design Research in the development of AI", in: Boess, Stella; Ming Cheung und Rebecca Cain (Hrsg.): *Proceedings of DRS 2020 International Conference*, S. 1315–1333, https://doi.org/10.21606/drs.2020.282.

Aviatar (2022): „How it works – What AVIATAR can do for you", https://www.aviatar.com/ how-it-works [zuletzt zugegriffen am 25.04.2022].

Baecker, Dirk (1999): „Organisation als System", Suhrkamp, Frankfurt.

Barocas, Solon und Helen Nissenbaum (2014): „Big Data's End Run around Anonymity and Consent", in: Lane, Julia; Victoria Stodden, Stefan Bender und Helen Nissenbaum (Hrsg.): „Privacy, Big Data, and the Public Good – Frameworks for Engagement", Cambridge University Press, S. 44–75, http://dx.doi.org/10.1017/CBO9781107590205. 004.

Barocas, Solon und Andrew D. Selbst (2016): „Big Data's Disparate Impact", *California Law Review*, Bd. 104, S. 671–732, http://dx.doi.org/10.2139/ssrn.2477899.

Barr, Alistair (2015): „Google Mistakenly Tags Black People as 'Gorillas', Showing Limits of Algorithms", The Wall Street Journal, https://www.wsj.com/articles/BL-DGB-42522 [zuletzt zugegriffen am 25.02.2025].

Bastani, Osbert, Carolyn Kim und Hamsa Bastani (2017): „Interpretability via model extraction", *Presented as a poster at the 2017 Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2017)*, S. 1–19, eprint arXiv:1706.09773.

Beauchamp, Tom und James Childress (1979): „Principles of biomedical ethics", Oxford University Press.

Beauchamp, Tom und James Childress (2019): „Principles of Biomedical Ethics: Marking Its Fortieth Anniversary", *The American Journal of Bioethics*, Bd. 19(11), S. 9–12, DOI: 10.1080/15265161.2019.1665402.

Benjamin, Ruha (2019): „Race After Technology: Abolitionist Tools for the New Jim Code", Polity Press, Cambridge, UK.

Beier, Kathi; Dagmar Borchers, Hans-Henrik Dassow, Martin Hähnel, Björn Haferkamp, Antonia Kempkens und Regina Müller (2024): „Tugendethische Ansätze", in: Grimm, Petra; Kai Erik Trost und Oliver Zöllner (Hrsg.): „Digitale Ethik", Nomos, Verlag Karl Alber, S. 49–62.

Belliger, Andréa und David J. Krieger (2023): „New Directions in Digital Ethics", in: Casas-Roma, Joan; Jordi Conesa und Santi Caballé (Hrsg.): „Technology, Users and Uses: Ethics and Human Interaction through Technology and AI", Ethics International Press, S. 1–19.

Bendig, Thomas; Philipp Ehmann, Pablo Mentzinis, Max Pfeifer, Jörg Beinert, Peter Ganten, Thomas Mosch, Thomas Schaf, Thomas Dapp, Pamela Krosta-Hartl, Rahild Neuburger und Fabian Zacharias (2019): „Digitale Souveränität im Kontext plattformbasierter Ökosysteme – Plattform ‚Innovative Digitalisierung der Wirtschaft', Fokusgruppe ‚Digitale Souveränität' im Rahmen des Digitalgipfels 2019", Bundesministerium für Wirtschaft und Energie.

Bentham, Jeremy (2017) [1789]: „An Introduction to the Principles of Morals and Legislation", kommentiert durch Jonathan Bennett, https://www.earlymoderntexts.com/assets/pdfs/bentham1780.pdf [zuletzt zugegriffen am 14.10.2022].

Berlin, Isaiah (1997): „The Proper Study of Mankind – An Anthology of Essays", Farrar, Straus and Giroux, New York.

Bernal, Paul (2016): „Data gathering, surveillance and human rights: recasting the debate", *Journal of Cyber Policy*, Bd. 1(2), S. 243–264, DOI: 10.1080/23738871.2016.1228990.

Betzler, Monika und Nina Scherrer (2016): „Verantwortung und Kontrolle", in: Heidbrink, Ludger; Claus Langbehn und Janina Sombetzki (Hrsg.): „Handbuch Verantwortung", Springer Reference Sozialwissenschaften, S. 1–16, DOI 10.1007/978–3–658–06175–3_48–1.

Birhane, Abeba und Vinay Uday Prabhu (2021): „Large Image Datasets: A Pyrrhic Win for Computer Vision?", *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, S. 1537–1547, https://doi.org/10.48550/arXiv.2006.16923.

Bleher, Hannah und Matthias Braun (2023): „Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice", *Science and Engineering Ethics*, Bd. 29(21), S. 1–23, https://doi.org/10.1007/s11948-023-00443-3.

Blok, Vincent und Pieter Lemmens (2015): „The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation", in: Koops, Bert-Jaap; Ilse Oosterlaken, Henny Romijn, Tsjalling Swierstra und Jeroen van den Hoven (Hrsg.): „Responsible Innovation 2 – Concepts, Approaches, and Applications", Springer International Publishing Switzerland, S. 19–35.

Boddington, Paula (2017): „Towards a code of ethics for artificial intelligence", Cham: Springer.

Boddington, Paula (2023): „AI Ethics – A Textbook", Springer Nature Singapore.

Bolte, Larissa und Aimee van Wynsberghe (2024): „Sustainable AI and the third wave of AI ethics: a structural turn", *AI Ethics*, Bd. 5, S. 1733–1742, https://doi.org/10.1007/s43681-024-00522-6.

Bonnemains, Vincent; Claire Saurel und Catherine Tessier (2018): „Embedded ethics: some technical and ethical challenges", *Ethics and Information Technology*, Bd. 20, S. 41–58, https://doi.org/10.1007/s10676-018-9444-x.

Boyd, Danah und Kate Crawford (2012): „Critical Questions for Big Data – Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication & Society*, Bd. 15(5), S. 662–679, DOI: 10.1080/1369118X.2012.678878.

Buijsman, Stefan (2022): „Defining Explanation and Explanatory Depth in XAI", *Minds and Machines*, Bd. 32, S. 563–584, https://doi.org/10.1007/s11023-022-09607-9.

Buijsman, Stefan (2023): „Why and How Should We Explain AI?", in: Chetouani, Mohamed; Virginia Dignum, Paul Lukowicz und Carles Sierra (Hrsg.): „Human-Centered Artificial Intelligence – Advanced Lectures", *Lecture Notes in Computer Science*, including subseries *Lecture Notes in Artificial Intelligence*, Bd. 13500, S. 196–215, Springer, https://doi.org/10.1007/978-3-031-24349-3_11.

Buijsman, Stefan und Herman Veluwenkamp (2023): „Spotting When Algorithms Are Wrong", *Minds and Machines*, Bd. 33, S. 541–562, https://doi.org/10.1007/s11023-022-09591-0.

Buijsman, Stefan und Juan M. Durán (2024): „Epistemic implications of machine learning models in science", in: Tarja Knuuttila, Rami Koskinen, Natalia Carrillo (Hrsg.): „The Routledge Handbook of Philosophy of Scientific Modeling", Taylor & Francis, S. 456–469, https://doi.org/10.4324/9781003205647-39.

Bundesministerium für Wirtschaft und Energie (2019): „Das Projekt GAIA-X – Eine vernetzte Dateninfrastruktur als Wiege eines vitalen, europäischen Ökosystems", https://www.bmwk.de/Redaktion/DE/Publikationen/Digitale-Welt/das-projekt-gaia-x.pdf?__blob=publicationFile&v=16 [zuletzt zugegriffen am 15.05.2025].

Buolamwini, Joy und Timnit Gebru (2018): „Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Proceedings of Machine Learning Research*, Bd. 81, Conference on Fairness, Accountability, and Transparency, S. 1–15.

Burckhart, Holger und Jürgen Nielsen-Sikora (2020): „Verantwortung", in: Weiß, Gabriele und Jörg Zirfas (Hrsg.): „Handbuch Bildungs- und Erziehungsphilosophie", Springer, S. 177–187, https://doi.org/10.1007/978-3-658-19004-0_16.

Burget, Mirjam; Emanuele Bardone und Margus Pedaste (2017): „Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review", *Science and Engineering Ethics*, Bd. 23, S. 1–19.

Burrell, Jenna (2016): „How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, Bd. 3(1), S. 1–12, DOI: 10.1177/2053951715622512.

Capurro, Rafael (2017): „Digitization as an ethical challenge", *AI & Society*, Bd. 32, S. 277–283, DOI 10.1007/s00146–016–0686-z.

Cavoukian, Ann (2012): „Privacy by Design", *IEEE Technology and Society Magazine,* Bd. 31(4), S. 18–19, doi: 10.1109/MTS.2012.2225459.

Cavoukian, Ann; Scott Taylor und Martin E. Abrams (2010): „Privacy by Design: essential for organizational accountability and strong business practices", *Identity in the Information Society*, Springer, Bd. 3, S. 405–413, https://doi.org/10.1007/s12394-010-0053-z.

Chen, Ying; Jeffrey Kreulen, Murray Campbell und Carl Abrams (2011): „Analytics Ecosystem Transformation: A force for business model innovation", Annual SRII Global Conference, *IEEE Computer Society*, DOI 10.1109/SRII.2011.12, S. 11–20.

Coeckelbergh, Mark (2016): „Responsibility and the Moral Phenomenology of Using Self-Driving Cars", *Applied Artificial Intelligence*, Bd. 30(8), S. 748–757. https://doi.org/10.1080/08839514.2016.1229759.

Colaner, Nathan (2022): „Is explainable artificial intelligence intrinsically valuable?", *AI & Society*, Bd. 37, S. 231–238, https://doi.org/10.1007/s00146-021-01184-2.

Comesaña, Juan (2010): „Evidentialist Reliabilism", *Noûs*, Wiley, Bd. 44(4), S. 571–600, https://doi.org/10.1111/j.1468-0068.2010.00748.x.

Conee, Earl and Richard Feldman (2004): „Evidentialism: Essays in Epistemology", Oxford University Press.

Cooley, Thomas McIntyre (1879): „ A Treatise on the Law of Torts or the Wrongs Which Arise Independent of Contract", Chicago: Callaghan and Company.

Cormen, Thomas H.; Charles E. Leiserson, Ronald Rivest und Clifford Stein (2013): „Algorithmen – Eine Einführung", 4. Auflage, Oldenbourg Verlag München.

Corsten, Hans und Stefan Roth (2022): „Gedanken zur Digitalisierung", in: Roth, Stefan und Hans Corsten (Hrsg.): „Handbuch Digitalisierung", Verlag Franz Vahlen München, S. 3–18, doi.org/10.15358/9783800665631.

Crawford, Kate (2024): „Generative AI is guzzling water and energy", *Nature*, Bd. 626, S. 693, doi: https://doi.org/10.1038/d41586-024-00478-x.

Danaher, John (2016): „Robots, law and the retribution gap", *Ethics and Information Technology*, Bd. 18, S. 299–309, DOI 10.1007/s10676–016–9403–3.

Danaher, John (2022): „Tragic Choices and the Virtue of Techno-Responsibility Gaps", *Philosophy & Technology*, Bd. 35(26), S. 1–26, https://doi.org/10.1007/s13347-022-00519-1.

Datta, Anupam; Shayak Sen und Yair Zick (2016): „Algorithmic transparency via quantitative input influence", *Proceedings of 37th IEEE symposium on security and privacy*, San Jose, USA, S. 1–20, https://www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf [zuletzt zugegriffen am 04.03.2025].

David, Marian (2001): „Truth and the Epistemic Goal", in: Steup, Matthias (Hrsg.): „Knowledge, Truth, and Duty – Essays on Epistemic Justification, Responsibility, and Virtue", New York: Oxford University Press, S. 151–169.

Davies, Tim (2011): „Open data: infrastructures and ecosystems", *Proceedings of web science conference*, S. 1–6.

Dawes, Sharon S.; Lyudmila Vidiasova und Olga Parkhimovich (2016): „Planning and designing open government data programs: an ecosystem approach", *Government Information Quarterly*, Bd. 33(1), S. 1–53, http://dx.doi.org/10.1016/j.giq.2016.01.003.

De Mesel, Benjamin und Sybren Heyndels (2019): „The Facts and Practices of Moral Responsibility", *Pacific Philosophical Quarterly*, Bd. 100(3), S. 790–811, https://doi.org/10.1111/papq.12276.

Dienlin, Tobias (2014): „The Privacy Process Model", in: Garnett, Simon; Stefan Halft, Matthias Herz, Julia Maria Mönig (Hrsg.): „Medien und Privatheit", Stutz, S. 105–122.

Dignum, Virginia; Matteo Baldoni, Cristina Baroglio, Maurizio Caon, Raja Chatila, Louise Dennis, Gonzalo Génova, Galit Haim, Malte S. Kließ, Maite Lopez-Sanchez, Roberto Micalizio, Juan Pavón, Marija Slavkovik, Matthijs Smakman, Marlies van Steenbergen, Stefano Tedeschi, Leon van der Torre, Serena Villata, Tristan de Wildt (2018): „Ethics by Design: Necessity or Curse?", *AIES '18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery*, S. 60–66, https://doi.org/10.1145/3278721.3278745.

Dodge, Jesse; Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A. Smith, Nicole DeCario und Will Buchanan (2022): „Measuring the carbon intensity of AI in cloud instances", *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, S. 1877–1894, https://doi.org/10.48550/arXiv.2206.05229.

Donges, Patrick (2022): „Digitalisierung der politischen Kommunikation. Folgen der automatisiert algorithmischen Vermittlung auf die Sichtbarkeit und Zurechenbarkeit des Politischen", *Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS)*, Bd. 74, S. 209–230, https://doi.org/10.1007/s11577-022-00834-7.

Dormandy, Katherine (2019): „Evidentialismus", in: Grajner, Martin und Guido Melchior (Hrsg.): „Handbuch Erkenntnistheorie", J.B. Metzler Verlag, S. 178–186.

Drew, Paul (1991): „Asymmetries of knowledge in conversational interactions", in: Markova, Ivana und Klaus Foppa (Hrsg.): „Asymmetries in dialogue", Hemel Hempstead: Harvester Wheatsheaf, S. 21–49.

Duden (2023): „Privat", https://www.duden.de/rechtschreibung/privat [zuletzt zugegriffen am 02.09.2025].

Duhigg, Charles (2012): „How Companies learn your secrets", New York Times Magazine, February 16th, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp [zuletzt zugegriffen am 02.09.2025].

Durán Juan Manuel und Nico Formanek (2018): „Grounds for trust: essential Epistemic opacity and computational Reliabilism", *Minds and Machines*, Bd. 28(4), S. 645–666, DOI 10.1007/s11023–018–9481–6.

Durán, Juan Manuel und Karin Rolanda Jongsma (2021): „Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI", *Journal of Medical Ethics*, Bd. 47(5), S. 329–335, https://doi.org/10.1136/medethics-2020-106820.

Durnell, Eric; Karynna Okabe-Miyamoto, Ryan T. Howell und Martin Zizi (2020): „Online Privacy Breaches, Offline Consequences: Construction and Validation of the
Concerns with the Protection of Informational Privacy Scale", *International Journal of Human–Computer Interaction,* Bd. 36(19), S. 1834–1848, https://doi.org/10.1080/10447318.2020.1794626

Ertel, Wolfgang (2021): „Grundkurs Künstliche Intelligenz – Eine praxisorientierte Einführung", 5. Auflage, Springer.

Espiner, Tom und Joe Tidy (2023): „Royal Mail hit by Russia-linked ransomware attack", BBC News, https://www.bbc.com/news/business-64244121 [zuletzt zugegriffen am 30.04. 2025].

Fahmy, Melissa Seymour (2023): „Never Merely as a Means: Rethinking the Role and Relevance of Consent", *Kantian Review*, Bd. 28, S. 41–62, doi:10.1017 /S136941542200053X.

Fantl, Jeremy und Matthew McGrath (2002): „Evidence, Pragmatics, and Justification", *The Philosophical Review*, Bd. 111(1), S. 67–94, DOI 10.1215/00318108–111–1–67.

Felder, Ekkehard (2013): „Faktizitätsherstellung mittels handlungsleitender Konzepte und agonaler Zentren – Der diskursive Wettkampf um Geltungsansprüche", in: Felder, Ekkehard (Hrsg.): „Faktizitätsherstellung in Diskursen: die Macht des Deklarativen", De Gruyter, S. 13–28, https://doi.org/10.1515/9783110289954.13.

Feng, Patrick (2000): „Rethinking Technology, Revitalizing Ethics: Overcoming Barriers to Ethical Design", *Science and Engineering Ethics*, Bd. 6, S. 207–220, https://doi.org/10.1007/s11948-000-0049-4.

Fischer, John Martin und Mark Ravizza (1998): „Responsibility and control. A theory of moral responsibility", 1. Auflage, Cambridge University Press.

Floridi, Luciano (2014): „Open Data, Data Protection, and Group Privacy", *Philosophy & Technology*, Bd. 27, S. 1–3, https://doi.org/10.1007/s13347-014-0157-8.

Floridi, Luciano und Mariarosaria Taddeo (2016): „What is data ethics?", *Philosophical Transactions*, The Royal Society, Bd. 374, S. 1–5, http://dx.doi.org/10.1098/rsta.2016.0360.

Floridi, Luciano (2017): „Group Privacy: A Defence and an Interpretation. *Group Privacy: New Challenges of Data Technologies*", in: Taylor, Linnet; Luciano Floridi, and Bart van der Sloot (Hrsg.), Springer, S. 83–100, https://doi.org/10.1007/978-3-319-46608-8_5.

Floridi, Luciano (2019): „Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical", *Philosophy & Technology*, Bd. 32, S. 185–193, https://doi.org/10.1007/s13347-019-00354-x.

Floridi, Luciano and Josh Cowls (2021): „A Unified Framework of Five Principles for AI in Society", in: Floridi, Luciano (Hrsg.): „Ethics, Governance, and Policies in Artificial Intelligence", Springer Nature, Philosophical Studies Series 144, https://doi.org/10.1007/978-3-030-81907-1_2.

Foley, Richard (1993): „Working without a net – A study of egocentric epistemology", Oxford University Press.

Foot, Philippa (2002) [1967]: „The Problem of Abortion and the Doctrine of the Double Effect", in: Foot, Philippa: „Virtues and Vices: and other essays in moral philosophy", Oxford University Press, S. 1–6.

Frankfurt, Harry G. (1971): „Freedom of the will and the concept of a Person", *The Journal of Philosophy*, Bd. 68(1), S. 5–20, https://doi.org/10.2307/2024717.

Fraser, Colin (2020): „Target didn't find out a teenager was pregnant before her father did, and that one article that said they did was silly and bad.", Medium, January 4th, https://medium.com/@colin.fraser/target-didnt-figure-out-a-teen-girl-was-pregnant-before-her-father-did-a6be13b973a5 [zuletzt zugegriffen am 02.09.2025].

Freiesleben, Timo und Gunnar König (2023): „Dear XAI Community, We Need to Talk! Fundamental Misconceptions in Current XAI Research", in: Longo, Luca (Hrsg.): „Explainable Artificial Intelligence", *Proceedings of the First World Conference*, Lisbon, Portugal, July 26–28, S. 1–17.

French, Peter A. (1979): „The corporation as a moral person", *American Philosophical Quarterly*, Bd. 16(3), S. 207–215, https://www.jstor.org/stable/20009760.

Fricker, Miranda (2016): „What's the Point of Blame? A Paradigm Based Explanation", *Noûs*, Wiley, Bd. 50(1), S. 165–183, https://doi.org/10.1111/nous.12067.

Fried, Charles (1968): „Privacy [A moral analysis]", in Schoeman, Ferdinand David (Hrsg.): „Philosophical Dimensions of Privacy: An Anthology", Cambridge University Press, S. 203- 223.

Fried, Charles (1970): „An Anatomy of Values", Harvard University Press.

Friedman, Batya (1997): „Human Values and the Design of Computer Technology", *CSLI Lecture Notes*, Bd. 72, New York: Cambridge University Press.

Friedman, Batya; Peter H. Kahn und Alan Borning (2002): „Value Sensitive Design: Theory and Methods", University of Washington technical report.

Friedman, Batya und Peter H. Kahn Jr. (2008): „Human values, ethics, and design", in: Sears, Andrew und Julie A. Jacko (Hrsg.): „The Human-Computer Interaction Handbook -Fundamentals, Evolving Technologies, and Emerging Applications", Second Edition, Taylor & Francis Group, S. 1241–1266.

Friedman, Batya; Peter H. Kahn und Alan Borning (2008): „Value sensitive design and information systems", in: Himma, Kenneth Einar und Herman T. Tavani (Hrsg.): „The Handbook of Information and Computer Ethics", John Wiley & Sons, Inc., S. 69–101, https://doi.org/10.1002/9780470281819.ch4.

Friedman, Batya und Helen Nissenbaum (1996): „Bias in Computer Systems", *ACM Transactions on Information Systems*, Bd. 14(3), S. 330–347, https://doi.org/10.1145/230538.23056.

Geiger, Christian P. und Jörn von Lucke (2012): „Open Government and (Linked) (Open) (Government) (Data)", *eJournal of eDemocracy and Open Government (JeDEM)*, Bd. 4(2), S. 265–278, DOI: 10.29379/jedem.v4i2.143.

Geuss, Raymond (2001): „Public Goods, Private Goods", Princeton University Press.

Ghanbari, Hadi und Kari Koskinen (2024): „When data breach hits a psychotherapy clinic: The Vastaamo case", *Journal of Information Technology*, Teaching Cases, Bd. 0(0), S. 1–9, DOI: 10.1177/20438869241258235.

Gogoll, Jan; Niina Zuber, Severin Kacianka, Timo Greger, Alexander Pretschner und Julian Nida-Rümelin (2021): „Ethics in the Software Development Process: from Codes of Conduct to Ethical Deliberation", *Philosophy & Technology*, Bd. 34, S. 1085–1108, https://doi.org/10.1007/s13347-021-00451-w.

Goldman, Alvin I. (1979): „What Is Justified Belief?", in: Pappas, Georges S. (Hrgs.): „Justification and Knowledge – New Studies in Epistemology", D. Reidel Publishing Company, S. 1–25, https://doi.org/10.1007/978-94-009-9493-5_1.

Görs, Philipp K.; Anne Traum, Arne Koevel und Friedemann W. Nerdinger (2022): „Wirkungen der Digitalisierung auf Mitarbeitende: Ein narrativer Review empirischer Befunde", in: Moukouli, Virginia; Friedemann W. Nerdinger, Hakan Yergün, Alrik Zech, Marco Zimmer (Hrsg.): „Kompetenzen von Mitarbeiterinnen und Mitarbeitern in der digitalisierten Arbeitswelt – Analysen und Handlungsempfehlungen am Beispiel der Steuerberatungsbranche", Springer Gabler, S. 11–32, https://doi.org/10.1007/978-3-658-34869-4_2.

Goodman, Bryce (2016): „Economic Models of (Algorithmic) Discrimination", *29th Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain, http://www.mlandthelaw.org/papers/goodman2.pdf [zuletzt zugegriffen am 24.02.2025].

Goodman, Bryce und Seth Flaxman (2017): „European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"", *AI Magazine,* Bd. 38(3), S. 50–57. https://doi.org/10.1609/aimag.v38i3.2741.

Granka, Laura A. (2010): „The Politics of Search: A Decade Retrospective", *The Information Society*, Bd. 26(5), S. 364–374, DOI: 10.1080/01972243.2010.511560.

Grosz, Barbara J.; David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons und Jim Waldo (2019): „Embedded EthiCS: Integrating Ethics Across CS Education", *Communications of the ACM*, Bd. 62(8), S. 54–61, https://doi.org/10.1145/3330794.

Grote, Thomas; Konstantin Genin und Emily Sullivan (2024): „Reliability in Machine Learning", *Philosophy Compass*, Bd. 19(5), e12974, S. 1–11, https://doi.org/10.1111/phc3.12974.

Guynn, Jessica (2015): „Google Photos Labeled Black People 'Gorillas' ", USA TODAY, http://www.usatoday.com/story/tech/2015/07/01/google-apologizesafter-photos-identify-black-people-as-gorillas/29567465 [zuletzt zugegriffen am 25.02.2025].

Ha, Suwook; Seungyun Lee und Kangchan Lee (2014): „Standardization Requirements Analysis on Big Data in Public Sector based on Potential Business Models", *International Journal of Software Engineering and Its Applications*, Bd. 8(11), S. 165–172, http://dx.doi.org/10.14257/ijseia.2014.8.11.15.

Hagendorff, Thilo (2020): „The ethics of AI Ethics: an evaluation of guidelines", *Minds and Machines*, Bd. 30, S. 99–120, https://doi.org/10.1007/s11023-020-09517-8.

Hallensleben, Sebastian; Carla Hustedt, Lajla Fetic, Torsten Fleischer, Paul Grünke, Thilo Hagendorff, Marc Hauer, Andreas Hauschke, Jessica Heesen, Michael Herrmann, Rafaela Hillerbrand, Christoph Hubig, Andreas Kaminski, Tobias Krafft, Wulf Loh, Philipp Otto und Michael Puntschuh (2020): „From Principles to Practice – An interdisciplinary framework to operationalise AI ethics", Manthey, Nora (Hrsg.), Artificial Intelligence Ethics Impact Group (AIEI), Bertelsmann Stiftung.

Haltaufderheide, Joschka (2015): „Zur Risikoethik. Analysen im Problemfeld zwischen Normativität und unsicherer Zukunft", Epistemata Philosophie, Königshausen & Neumann, Bd. 565, S. 1–296.

Hansson, Sven Ove (1999): „A Philosophical Perspective on Risk", *Ambio*, Bd. 28(6), S. 539–542, https://www.jstor.org/stable/4314951.

Hansson, Sven Ove (2003): „Ethical Criteria of Risk Acceptance", *Erkenntnis*, Bd. 59(3) S. 291–309, https://doi.org/10.1023/A:1026005915919.

Hansson, Sven Ove (2004): „Philosophical Perspectives on Risk", *Techné*, Bd. 8(1), S. 10–35, DOI 10.5840/techne2004818.

Hansson, Sven Ove (2005): „Seven Myths of Risk", *Risk Management: An International Journal*, Bd. 7 (2), S. 7–17, https://doi.org/10.1057/palgrave.rm.8240209.

Hansson, Sven Ove (2018): „How to Perform an Ethical Risk Analysis (eRA)", *Risk Analysis*, Bd. 38(9), S. 1820–1829, DOI: 10.1111/risa.12978.

Hardwig, John (1985): „Epistemic Dependence", *The Journal of Philosophy*, Bd. 82(7), S. 335–349, https://doi.org/10.2307/2026523.

Hare, Richard M. (1993): „Could Kant Have been A Utilitarian?", *Utilitas*, Bd. 5(1), S. 1–16.

Harris, Keith (2022): „Epistemic Domination", *Thought: A Journal of Philosophy*, Bd. 11(3), S. 134–141, https://doi.org/10.5840/tht202341317.

Harrison, Teresa M.; Theresa A. Pardo und Meghan Cook (2012): „Creating open government ecosystems: a research and development agenda", *Future Internet*, Bd. 4, S. 900–928, doi:10.3390/fi4040900.

Hart, Herbert Lionel Adolphus (1968): „Punishment and Responsibility. Essays in the Philosophy of Law", Oxford: Clarendon Press.

Harwardt, Mark (2022): „Management der digitalen Transformation – Eine praxisorientierte Einführung", 2. Auflage, Springer Gabler.

Hawthorne, John und Jason Stanley (2008): „Knowledge and action", *Journal of Philosophy*, Bd. 105(10), S. 571–590, https://www.jstor.org/stable/20620129.

Heimstädt, Maximilian; Fredric Saunderson und Tom Heath (2014): „Conceptualizing open data ecosystems: a timeline analysis of open data development in the UK", Working Paper, Diskussionsbeiträge No. 2014/12, Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft.

Helsingin Sanomat (2024): „Vastaamo-uhrien juristi: Ihmisiä on päätynyt itsemurhaan tietomurron ja kiristyksen takia", https://www.hs.fi/suomi/art-2000010265660.html [zuletzt zugegriffen am 02.09.2025]..

Hermansson, Hélène und Sven Ove Hansson (2007): „A three-party model tool for ethical risk analysis", *Risk Management*, Bd. 9, S. 129–144, DOI: 10.1057/palgrave.rm.8250028.

Herzog, Christian (2021): „On the risk of confusing interpretability with explicability", *AI and Ethics*, Bd. 2, S. 219–225, https://doi.org/10.1007/s43681-021-00121-9.

Herzog, Christian (2022): „On the Ethical and Epistemological Utility of Explicable AI in Medicine", *Philosophy & Technology*, Bd. 35, S. 1–31, https://doi.org/10.1007/s13347-022-00546-y.

Hevelke, Alexander und Julian Nida-Rümelin (2015): „Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis", *Science and Engineering Ethics*, Bd. 21, S. 619–630, DOI 10.1007/s11948–014–9565–5.

Hildebrandt, Mireille (2020): „Law for Computer Scientists and Other Folk", Oxford University Press.

Hill, Kashmir (2012): „How Target figured out a teen girl was pregnant before her father did", Forbes, February 16th, https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/ [zuletzt zugegriffen am 02.09.2025]..

Hill, Thomas E. (2002): „Human Welfare and Moral Worth: Kantian Perspectives", Oxford University Press.

Himmelreich, Johannes (2019): „Responsibility for Killer Robots", *Ethical Theory and Moral Practice*, Bd. 22, S. 731–747, https://doi.org/10.1007/s10677-019-10007-9.

Himmelreich, Johannes und Sebastian Köhler (2022): „Responsible AI Through Conceptual Engineering", *Philosophy & Technology*, Bd. 35(60), S. 1–30, https://doi.org/10.1007/s13347-022-00542-2.

Holter, Carolyn Ten (2022): „Participatory design: lessons and directions for responsible research and innovation", *Journal of Responsible Innovation*, Bd. 9(2), S. 275–290, DOI: 10.1080/23299460.2022.2041801.

Holzinger, Andreas; Anna Saranti, Christoph Molnar, Przemyslaw Biecek und Wojciech Samek (2022): „Explainable AI Methods – A Brief Overview", in: Holzinger, Andreas; Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller und Wojciech Samek (Hrsg.): „xxAI – Beyond Explainable AI", *Lecture Notes in Computer Science*, Bd. 13200. Springer, S. 13–38, https://doi.org/10.1007/978-3-031-04083-2_2.

Hume, David (2000) [1739]: „A Treatise of Human Nature", David Fate Norton and Mary J. Norton (Hrsg.), Oxford Philosophical Texts.

Humphreys, Paul (2009): „The philosophical novelty of computer simulation methods", *Synthese*, Bd. 169(3), S. 615–626, https://doi.org/10.1007/ s11229–008–9435–2.

Iansiti, Marco und Roy Levien (2004): „Strategy as Ecology", Harvard Business Review, S. 68–78, https://hbr.org/2004/03/strategy-as-ecology [zuletzt zugegriffen am 12.07.2022].

Ichikawa, Jonathan Jenkins and Matthias Steup (2018): „The Analysis of Knowledge", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (Hrsg.), https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/ [zuletzt zugegriffen am 11.10. 2023].

Immonen, Anne; Marko Palviainen und Eila Ovaska (2014): „Requirements of an Open Data Based Business Ecosystem", *IEEE Access*, Bd. 2, S. 88–103, doi: 10.1109/ACCESS.2014.2302872.

Inness, Julie (1992): „Privacy, Intimacy, and Isolation", Oxford University Press.

Isaacs, Tracy (2016): „Kollektive Verantwortung", in: Heidbrink, Ludger; Claus Langbehn und Janina Sombetzki (Hrsg.): „Handbuch Verantwortung", Springer Reference Sozialwissenschaften, S. 1–23, DOI 10.1007/978–3–658–06175–3_25–1.

Jakobsen, Stig-Erik; Arnt Fløysand und John Overton (2019): „Expanding the field of Responsible Research and Innovation (RRI) – from responsible research to responsible innovation", *European Planning Studies*, Bd. 27(12), S. 2329–2343, DOI: 10.1080/09654313.2019.1667617.

Jobin, Anna; Marcello Ienca und Effy Vayena (2019): „The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Bd. 1, S. 389–399, https://doi.org/10.1038/s42256-019-0088-2.

Jonas, Hans (2020) [1979]: „Das Prinzip Verantwortung – Versuch einer Ethik für die technologische Zivilisation ", Suhrkamp Verlag.

Kamm, Frances M. (2020): „The Use and Abuse of the Trolley Problem Self-Driving Cars, Medical Treatments, and the Distribution of Harm", in: Liao, Matthew S.: „Ethics of Artificial Intelligence", Oxford University Press, S. 79–108, DOI: 10.1093/oso/9780190905040. 003.0003.

Kane, Robert (1996): „The significance of free will", Oxford University Press.

Kant, Immanuel (1784): „Beantwortung der Frage: Was ist Aufklärung?", *Berlinische Monatsschrift*, Dezember-Heft, S. 481–494.

Kant, Immanuel (1956): „Kritik der reinen Vernunft", Philosophische Bibliothek, Band 37a), Raymund Schmidt (Hrsg.).

Kant, Immanuel (2008): „Grundlegung zur Metaphysik der Sitten", Philipp Reclam jun. GmbH & Co. KG, Stuttgart.

Kellmeyer, Philipp (2024): „Beyond participation: Towards a community-led approach to value alignment of AI in medicine", *Developments in Neuroethics and Bioethics*, Bd. 7, S. 249–269, https://doi.org/10.1016/bs.dnb.2024.02.011.

Kempkens, Antonia (2025): „Ethical Design of Datafication by Principles of Biomedical Ethics", in: Hähnel, Martin und Regina Müller (Hrsg.): „A Companion to Applied Philosophy of AI", Wiley Blackwell, S. 224 – 237, DOI: 10.1002/9781394238651.ch16.

Kerstein, Samuel (2009): „Treating Others Merely as Means", *Utilitas*, Bd. 21(2), S. 163–180, doi:10.1017/S0953820809003458.

Kerstein, Samuel (2013): „How to treat persons", Oxford University Press.

Keymolen, Esther und Linnet Taylor (2023): „Data Ethics and Data Science: An Uneasy Marriage?“, in: Liebregts, Werner; Willem-Jan van den Heuvel und Arjan van den Born (Hrsg.): „Data Science for Entrepreneurship – Principles and Methods for Data Engineering, Analytics, Entrepreneurship, and the Society”, Classroom Companion: Business, Springer, S. 481–499, https://doi.org/10.1007/978-3-031-19554-9_20.

Kiener, Maximilian (2022): „Can we Bridge AI’s responsibility gap at Will?“, *Ethical Theory and Moral Practice*, Bd. 25, S. 575–593, https://doi.org/10.1007/s10677-022-10313-9.

Kiran, Asle H.; Nelly Oudshoorn und Peter-Paul Verbeek (2015): „Beyond checklists: toward an ethical-constructive technology assessment“, *Journal of Responsible Innovation*, Bd. 2(1), S. 5–19, DOI: 10.1080/23299460.2014.992769.

Kitchin, Rob (2016): „Thinking critically about and researching algorithms“, *Information, Communication & Society*, Bd. 20(1), S. 14–29, https://doi.org/10.1080/1369118X.2016.1154087.

Kleingeld, Pauline (2024): „Wie man jemanden „bloß als Mittel“ gebraucht“, *Zeitschrift für Ethik und Moralphilosophie*, Bd. 7, S. 107–130, https://doi.org/10.1007/s42048-024-00175-4.

Klenk, Tanja; Frank Nullmeier und Göttrik Wewer (2020): „Auf dem Weg zum Digitalen Staat? Stand und Perspektiven der Digitalisierung in Staat und Verwaltung“, in: Klenk, Tanja; Frank Nullmeier und Göttrik Wewer (Hrsg.): „Handbuch Digitalisierung in Staat und Verwaltung“, Springer VS, S. 3–23.

Knight, Frank H. (1921): „Risk, Uncertainty, and Profit“, Houghton Mifflin Company, The Riverside Press Cambridge.

Koch, Steffen (2019): „Reliabilismus”, in: Grajner, Martin und Guido Melchior (Hrsg.): „Handbuch Erkenntnistheorie“, J.B. Metzler Verlag, S. 169–177, DOI: 10.1007/978–3–476–04632–1_22.

Köver, Chris (2019): „Streit um den AMS-Algorithmus geht in die nächste Runde“, Netzpolitik.org, https://netzpolitik.org/2019/streit-um-den-ams-algorithmus-geht-in-die-naechste-runde/ [zuletzt zugegriffen am 05.03.2025].

Kolkman, Daan (2022): „The (in)credibility of algorithmic models to non-experts“, *Information, Communication & Society*, Bd. 25(1), S. 93–109, DOI: 10.1080/1369118X.2020.1761860.

Korsgaard, Christine (1996): „Creating the Kingdom of Ends”, Cambridge University Press.

Krafft, Tobias D. und Katharina Zweig (2019): „Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse – Ein Regulierungsvorschlag auch sozioinformatischer Perspektive", Verbraucherzentrale Bundesverband e.V., https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf [zuletzt zugegriffen am 04.04.2024].

Krainer, Larissa (2024): „Diskriminierung", in: Grimm, Petra; Kai Erik Trost und Oliver Zöllner (Hrsg.): „Digitale Ethik", Nomos, Verlag Karl Alber, S. 303–314.

Krieger, David J. und Andréa Belliger (2014): „Interpreting Networks – Hermeneutics, Actor-Network Theory & New Media", transcript Verlag.

Krishnan, Maya (2020): „Against interpretability: a critical examination of the interpretability problem in machine learning" *Philosophy & Technology*, Bd. 33(3), S. 487–502, https://doi.org/10.1007/s13347-019-00372-9.

Kutz, Christopher (2000): „Complicity: Ethics and law for a collective age", Cambridge University Press.

Laas, Kelly; Michael Davis und Elisabeth Hildt (2022): „An Introduction to the Societal Roles of Ethics Codes", in: Laas, Kelly; Michael Davis, and Elisabeth Hildt (Hrsg.): "Codes of Ethics and Ethical Guidelines – Emerging Technologies, Changing Fields", *The International Library of Ethics, Law and Technology*, Bd. 23, S. 1–13, DOI: 10.1007/978–3–030–86201–5_1.

Lämmel, Uwe und Jürgen Cleve (2008): „Künstliche Intelligenz", 3. Auflage, Carl Hanser Verlag, München.

Lauer, Dave (2021): „You cannot have AI ethics without ethics", *AI and Ethics*, Bd. 1, S. 21–25, https://doi.org/10.1007/s43681-020-00013-4.

Lehmann, Helko; Dirk Werth, Christian Weber, Peter Henning, Christoph Meinel, Lars Nagel, Rainer Sträter, Heike Antvogel, Thomas Hoppe, Thomas Lachmann, Kerstin-Alexandra Bootsmann (2021): „Gesamtvorhabenbeschreibung: Förderwettbewerb „Innovative und praxisnahe Anwendungen und Datenräume im digitalen Ökosystem GAIA-X". MERLOT – MarkEtplace foR LifelOng educaTional dataspaces and smart service provisioning", [interner Zugang durch Dataport].

Lincke, Susan J. (2016): „Integrating Ethics and Risk Management", *4th International Symposium on Digital Forensic and Security (ISDFS)*, IEEE, S. 78–83, doi: 10.1109/ISDFS.2016.7473522.

Lindemann, Hilde (2019): „An Invitation to Feminist Ethics (2nd edn)", Oxford University Press.

Lindman, Juho; Tomi Kinnari und Matti Rossi (2015): „Business Roles in the Emerging Open-Data Ecosystem", *IEEE Software*, IEEE Computer Society, Bd. 33(5), S. 54–59, DOI: 10.1109/MS.2015.25.

Lipton, Zachary C. (2018): „The Mythos of Model Interpretability", *Queue*, Bd. 16(3), S. 31–57, https://doi.org/10.48550/arXiv.1606.03490.

Littlejohn, Clayton (2009): „Must we act only on what we know?", *Journal of Philosophy*, Bd. 106(8), S. 463–473, https://www.jstor.org/stable/20620191.

Looi, Jeffrey CL; Stephen Allison, Tarun Bastiampillai, Paul A Maguire, Steve Kisely, Sharon Reutens, Richard CH Looi (2025): „Cybersecurity lessons from the Vastaamo psychotherapy data breach for psychiatrists and other mental healthcare providers", *Australasian Psychiatry*, Bd. 33(1), S. 106–110, DOI: 10.1177/10398562241291340.

London, Alex John (2019): „Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability", *Hastings Center Report*, Bd. 49(1), S. 15–21, doi: 10.1002/hast.973.

Lubberink, Rob; Vincent Blok, Johan van Ophem und Onno Omta (2017): „Lessons for Responsible Innovation in the Business Context: A Systematic Literature Review of Responsible, Social and Sustainable Innovation Practices", in: Brem, Alexander; Bernd Stahl, Doris Schroeder, André Martinuzzi und Vincent Blok (Hrsg.): Special Issue „Responsible Research and Innovation (RRI) in Industry", *Sustainability*, Bd. 9(5), S. 1–31, https://doi.org/10.3390/su9050721.

Luhmann, Niklas (1990): „Risiko und Gefahr", in: Luhmann, Niklas (Hrsg.): „Soziologische Aufklärung 5 – Konstruktivistische Perspektiven", Westdeutscher Verlag, S. 131–169.

Lundgren, Björn (2020): „A Dilemma for Privacy as Control", *The Journal of Ethics*, Bd. 24, S. 165–175, https://doi.org/10.1007/s10892-019-09316-z.

Lundgren, Björn (2021a): „Confusion and the Role of Intuitions in the Debate on the Conception of the Right to Privacy", *Res Publica*, Bd. 27, S. 669–674, https://doi.org/10.1007/s11158-020-09495-9.

Lundgren, Björn (2021b): „How can we make sense of control-based intuitions for limited access conceptions of the right to privacy", *Journal of Ethics and Social Philosophy*, Bd. 20(3), S. 382–391, https://doi.org/10.26556/jesp.v20i3.1438.

Macnish, Kevin (2018): „Government Surveillance and Why Defining Privacy Matters in a Post-Snowden World", *Journal of Applied Philosophy*, Bd. 35(2), S. 417–432, doi: 10.1111/japp.12219.

Macnish, Kevin (2020): „Mass Surveillance: A Private Affair?", *Moral Philosophy and Politics*, Bd. 7(1), S. 9–27, https://doi.org/10.1515/mopp-2019-0025.

Mainz, Jakob (2022): „Inferences and the Right to Privacy", *The Journal of Value Inquiry*, Bd. 58, S. 563–581, https://doi.org/10.1007/s10790-022-09911-8.

Mainz, Jakob Thrane und Rasmus Uhrenfeldt (2021): „Too Much Info: Data Surveillance and Reasons to Favor the Control Account of the Right to Privacy", *Res Publica*, Bd. 27, S. 287–302, https://doi.org/10.1007/s11158-020-09473-1.

Manovich, Lev (2012): „Trending: The Promises and the Challenges of Big Social Data", in: Gold, Matthew K. (Hrsg.): "Debates in the Digital Humanities", University of Minnesota Press, S. 460–475, https://doi.org/10.5749/minnesota/9780816677948.003.0047.

Manson, Neil C. und Onora O'Neill (2007): „Rethinking informed consent in bioethics", Cambridge University Press.

Manzeschke, Arne (2021): „Digitalisierung und Organisationsethik. Ethische und technikphilosophische Skizzen", *Ethik in der Medizin*, Bd. 33, S. 219–232, https://doi.org/10.1007/s00481-021-00630-5.

Martani, Andrea und Patrik Hummel (2022): „Eine neue Generation des Datenschutzes? Gegenwärtige Unvollständigkeit, mögliche Lösungswege und nächste Schritte", in: Richter, Gesine; Wulf Loh, Alena Buyx und Sebastian Graf von Kielmansegg (Hrsg.): „Datenreiche Medizin und das Problem der Einwilligung, Ethische, rechtliche und sozialwissenschaftliche Perspektiven", Springer, S. 27–54, https://doi.org/10.1007/978-3-662-62987-1_3.

Masur, Philipp K; Doris Teutsch und Tobias Dienlin (2018): „Privatheit in der Online-Kommunikation", in: Schweiger, Wolfgang, Klaus Beck und Veronika Karnowski (Hrsg.): „Handbuch Online-Kommunikation", Springer Reference Sozialwissenschaften, S. 1–29, https://doi.org/10.1007/978-3-658-18017-1_16-1.

Matthias, Andreas (2004): „The responsibility gap: Ascribing responsibility for the actions of learning automata", *Ethics and Information Technology*, Bd. 6, S. 175–183, https://doi.org/10.1007/s10676-004-3422-1.

McCarthy, John; Marvin Minsky, Nathaniel Rochester und Claude Elwood Shannon (1955): „A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf [zuletzt zugegriffen am 07.07.2023].

McLennan, Stuart; Amelia Fiske, Leo Anthony Celi, Ruth Müller, Jan Harder, Konstantin Ritt, Sami Haddadin and Alena Buyx (2020): „An embedded ethics approach for AI development", *Nature Machine Intelligence*, Bd. 2, S. 488–490, https://doi.org/10.1038/s42256-020-0214-1.

McLennan, Stuart; Amelia Fiske, Daniel Tigard, Ruth Müller, Sami Haddadin and Alena Buyx (2022): „Embedded ethics: a proposal for integrating ethics into the development of medical AI", *BMC Medical Ethics*, Bd. 23(6), S. 1–10, https://doi.org/10.1186/s12910-022-00746-3.

McNamara, Andrew, Justin Smith and Emerson Murphy-Hill (2018): „Does ACM's code of ethics change ethical decision making in software development?", *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (ESEC/FSE '18), S. 729–733, https://doi.org/10.1145/3236024.3264833.

Menges, Leonhard (2020a): „Did the NSA and GCHQ Diminish Our Privacy? What the Control Account Should Say", *Moral Philosophy and Politics*, Bd. 7(1), S. 29–48, https://doi.org/10.1515/mopp-2019-0063.

Menges, Leonhard (2020b): „Responsibility and appropriate blame: The no difference view", *European Journal of Philosophy*, Bd. 29(2), S. 393–409, https://doi.org/10.1111/ejop.12571.

Menges, Leonhard (2020c): „A Defense of Privacy as Control", *The Journal of Ethics*, Bd. 25, S. 385–402, https://doi.org/10.1007/s10892-020-09351-1.

Menges, Leonhard (2022): „Three Control Views on Privacy", *Social Theory and Practice*, Bd. 48(4), S. 691–711, https://doi.org/10.5840/soctheorpract2022811172.

Merrill, Jeremy B. (2016): „Liberal, Moderate or Conservative? See How Facebook Labels You", The New York Times, August 23rd, https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html [zuletzt zugegriffen am 02.09.2025].

Mieth, Corinna und Christoph Bambauer (2016): „Verantwortung und Pflichten", in: Heidbrink, Ludger; Claus Langbehn und Janina Sombetzki (Hrsg.): „Handbuch Verantwortung", Springer Reference Sozialwissenschaften, S. 1–12, DOI 10.1007/978–3–658–06175–3_48–1.

Minsky, Marvin (1968): „Semantic Information Processing", Cambridge, MA: MIT Press.

Mitchell, Tom M. (1997): „Machine Learning", McGraw-Hill Science/Engineering/Math.

Mittelstadt, Brent (2019): „Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, Bd. S. 501–507, DOI: 10.1038/s42256–019–0114–4.

Mittelstadt, Brent Daniel; Patrick Allo, Mariarosaria Taddeo, Sandra Wachter und Luciano Floridi (2016): „The ethics of algorithms: Mapping the debate", *Big Data & Society*, Bd. 3(2), S. 1–21, https://doi.org/10.1177/2053951716679679.

Möller, Niklas (2009): „Should we follow the experts' advice? Epistemic uncertainty, consequence dominance and the knowledge asymmetry of safety", *International Journal of Risk Assessment and Management*, Bd. 11(3/4), S. 219–236, https://doi.org/10.1504/IJRAM.2009.023154.

Möller, Niklas; Sven Ove Hansson und Martin Peterson (2006): „Safety is more than the antonym of risk", *Journal of Applied Philosophy*, Bd. 23(4), S. 419–432, https://doi.org/10.1111/j.1468-5930.2006.00345.x.

Mohr, Georg (2010): „Moralische Rechte gibt es nicht", in: Sandkühler, Hans Jörg (Hrsg.): „Recht und Moral", Felix Meiner Verlag Hamburg.

Molnar, Christoph; Gunnar König, Julia Herbinger, Timo Freiesleben, Susanne Dandl, Christian A Scholbeck, Giuseppe Casalicchio, Moritz Grosse-Wentrup, Bernd Bischl (2022): „General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models", in: Holzinger, Andreas; Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller und Wojciech Samek (Hrsg.): „xxAI – Beyond Explainable AI", Lecture Notes in Computer Science, Bd. 13200. Springer, S. 39–68, https://doi.org/10.1007/978-3-031-04083-2_4.

Moore, Adam D. (2010): „Privacy Rights: Moral and Legal Foundations", Pennsylvania State University Press.

Morley, Jessica; Luciano Floridi, Libby Kinsey und Anat Elhalal (2020): „From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices", *Science and Engineering Ethics*, Bd. 26, S. 2141–2168, https://doi.org/10.1007/s11948-019-00165-5.

Morley, Jessica; Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander und Luciano Floridi (2021): „Ethics as a Service: A Pragmatic Operationalisation of AI Ethics", *Minds and Machines*, Bd. 31, S. 239–256, https://doi.org/10.1007/s11023-021-09563-w.

Müller, Vincent C. (2023): „The History of Digital Ethics", in: Véliz, Carissa (Hrsg.): "Oxford Handbook of Digital Ethics", Oxford University Press, S. 3–19, https://doi.org/10.1093/oxfordhb/9780198857815.013.1

Mulvenna, Maurice; Jennifer Boger und Raymond Bond (2017): „Ethical by Design: A Manifesto", *Proceedings of European Conference on Cognitive Ergonomics 2017* (ECCE 2017), Umeå, Sweden, September 19–22, 2017, S. 51–54, https://doi.org/10.1145/3121283.3121300.

Munch, Lauritz und Jakob Mainz (2023): „To Believe, or Not to Believe – That is Not the (Only) Question: The Hybrid View of Privacy", *The Journal of Ethics*, Bd. 27, S. 245–261, https://doi.org/10.1007/s10892-023-09419-8.

Munch, Lauritz Aastrup; Jens Christian Bjerring und Jakob Thrane Mainz (2024): „Algorithmic decision-making: The right to explanation and the significance of stakes", *Big Data & Society*, Bd. 11(1), S. 1–12, DOI: 10.1177/20539517231222872.

Munch, Lauritz; Jakob Mainz und Jens Christian Bjerring (2023): „The value of responsibility gaps in algorithmic decision-making", *Ethics and Information Technology*, Bd. 25(21), S. 1–11, https://doi.org/10.1007/s10676-023-09699-6.

Munn, Luke (2023): „The uselessness of AI ethics", *AI and Ethics*, Bd. 3, S. 869–877, https://doi.org/10.1007/s43681-022-00209-w.

Murphy, Matt A. (2023): „Using structured ethical techniques to facilitate reasoning in technology ethics", *AI and Ethics*, Bd. 5, S. 479–488, https://doi.org/10.1007/s43681-023-00371-9.

Nagel, Lars und Douwe Lycklama (2021): „Design Principles for Data Spaces", Position Paper, International Data Spaces Association, DOI: 10.18154/RWTH-2022–09082.

Nardi, Bonnie A. und Vicki L. O'Day (1999): „Information Ecologies: Using Technology with Heart", MIT Press.

Narveson, Jan (2002): „Collective responsibility", *The Journal of Ethics*, Bd. 6, S. 179–198, https://doi.org/10.1023/A:1015823716891.

Neumann, Peter G. (2009): „Computer-Related Risk Futures", *Proceedings of the 2009 Annual Computer Security Applications Conference*, IEEE Computer Society, S. 1–6, https://doi.org/10.1109/ACSAC.2009.13.

Niehaves, Björn and Oliver Heger (2015): „Verantwortungsvoll gestalten", in Habscheid, Stephan, Gero Hoch, Hilde Schröteler-von Brandt und Volker Stein (Hg.): „Gestalten gestalten", V&R unipress, S. 169–179.

Nissenbaum, Helen (2010): „Privacy in Context – Technology, Policy and the Integrity of Social Life", Stanford University Press.

Nyholm, Sven (2018): „Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and ResponsibilityLoci", *Science and Engineering Ethics*, Bd. 24, S. 1201–1219, https://doi.org/10.1007/s11948-017-9943-x.

Nyholm, Sven (2024): „Responsibility Gaps, Value Alignment, and Meaningful Human Control over Artificial Intelligence", in: Placani, Adriana und Stearns Broadhead (Hrsg.): „Risk and Responsibility in Context", Routledge, S. 191–213.

Oberdiek, John (2017): „Imposing Risk: A Normative Framework", Oxford University Press.

Öhman, Carl und David Watson (2019): „Digital Ethics: Goals and Approach", in: Öhman, Carl und David Watson (Hrsg.): „The 2018 Yearbook of the Digital Ethics Lab", Springer Nature Switzerland, S. 1–8, https://doi.org/10.1007/978-3-030-17152-0_1.

Oetzel, Marie Caroline und Sarah Spiekermann (2014): „A systematic methodology for privacy impact assessments: A design science approach", *European Journal of Information Systems*, Bd. 23(2), S. 126–150, https://doi.org/10.1057/ejis.2013.18.

Oimann, Ann-Katrien und Fabio Tollon (2024): „Responsibility Gaps and Technology: Old Wine in New Bottles?", *Journal of Applied Philosophy*, Bd. 42(1), S. 337–356, doi: 10.1111/japp.12763.

Oliveira, Marcelo Iury S. und Bernadette Farias Lóscio (2018): „What is a Data Ecosystem?", in: Anneke Zuiderwijk and Charles C. Hinnant (Hrsg.): „Proceedings of the 19th Annual International Conference on Digital Government Research", Association for Computing Machinery, S. 1–9, https://doi.org/10.1145/3209281.3209335.

Oliveira, Marcelo Iury S.; Glória de Fátima Barros Lima und Bernadette Farias Lóscio (2019): „Investigations into Data Ecosystems: a systematic mapping study", *Knowledge and Information Systems*, Bd. 61, S. 589–630, https://doi.org/10.1007/s10115-018-1323-6.

O'Neill, Onora (1989): „Constructions of Reason: Explorations of Kant's Practical Philosophy", Cambridge University Press.

O'Neill, Onora (2021): „A philosopher looks at digitale communication", Cambridge University Press.

Open Knowledge Foundation (2022): „Open Definition – Defining Open in Open Data, Open Content and Open Knowledge", https://opendefinition.org/ [zuletzt zugegriffen am 05.07.2022].

Owen, Richard; Jack Stilgoe, Phil Macnaghten, Mike Gorman, Erik Fisher and Dave Guston (2013): „A Framework for Responsible Innovation", in: Owen, Richard; John Bessant and Maggy Heintz (Hrsg.): „Responsible Innovation – Managing the responsible emergence of science and innovation in society", Wiley, S. 27–50.

Papadaki, Lina (2016): „Treating Others Merely as Means: A Reply to Kerstein", *Utilitas*, Bd. 28(1), S. 73–100, doi:10.1017/S0953820815000138.

Papanek, Victor (2009) [1984]: „Design für die reale Welt – Anleitung für eine humane Ökologie und sozialen Wandel", Pumhösl, Florian; Thomas Geisler, Martina Fineder und Gerald Bast (Hrsg.), Wien/New York: Springer.

Parent, William A. (1983a): „Privacy, Morality, and the Law", *Philosophy & Public Affairs*, Bd. 12(4), S. 269–288, https://www.jstor.org/stable/2265374.

Parent, William A. (1983b): „Recent Work on the Concept of Privacy", *American Philosophical Quarterly*, Bd. 20(4), S. 341–355.

Parsons, Mark A.; Øystein Godøy, Ellsworth LeDrew, Taco F. de Bruin, Bruno Danis, Scott Tomlinson und David Carlson (2011): „A conceptual framework for managing very diverse data for complex, interdisciplinary science," *Journal of Information Science*, Bd. 37(6), S. 555–569, https://doi.org/10.1177/0165551511412705.

Pasquale, Frank (2015): „The Black Box Society: The Secret Algorithms That Control Money and Information", Harvard University Press.

Persson, Ingmar und Julian Savulescu (2022): „The Impossibility of a Moral Right to Privacy", *Neuroethics*, Bd. 15(23), S. 1–5, https://doi.org/10.1007/s12152-022-09500-3.

Pflügl, Jakob (2024): „Entscheidet die KI über Jobs? Höchstgericht lässt AMS-Algorithmus erneut prüfen", Der Standard, https://www.derstandard.de/story/3000000206845/entscheidet-die-ki-ueber-jobs-hoechstgericht-laesst-ams-algorithmus-erneut-pruefen [zuletzt zugegriffen am 05.03.2025].

Poikola, Antti; Petri Kola und Kari A. Hintikka (2011): „Public data – an introduction to opening information resources", Ministry of Transport and Communications, Finnland.

Pollock, Rufus (2011): „Building the (open) data ecosystem", Open knowledge foundation Blog 31. https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/ [zuletzt zugegriffen am 04.07.2022].

Price, W. Nicholson und I. Glenn Cohen (2019): „Privacy in the age of medical big data", *Nature Medicine*, Bd. 25, S. 37–43, https://doi.org/10.1038/s41591-018-0272-7.

Rachels, James (1975): „Why Privacy Is Important", *Philosophy & Public Affairs*, Bd. 4 (4), S. 323–333, https://www.jstor.org/stable/2265077.

Rafner, Janet; Dominik Dellermann, Arthur Hjorth, Dóra Verasztó, Constance Kampf, Wendy Mackay und Jacob Sherson (2021): „Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence", *Morals + Machines*, Bd 2, S. 25–39, https://doi.org/ 10.5771/2747–5174–2021–2–24.

Raji, Inioluwa Deborah; Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron und Parker Barnes (2020): „Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing", *FAT'20: Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery (ACM)*, S. 33–44, https://doi.org/10.1145/3351095.3372873.

Rawls, John (1975): „Eine Theorie der Gerechtigkeit", Suhrkamp Taschenbuch Wissenschaft.

Reilly, Michael (2017): „Is Facebook Targeting Ads at Sad Teens?", *MIT Technology Review*, May 1st, https://www.technologyreview.com/2017/05/01/105987/is-facebook-targeting-ads-at-sad-teens/ [zuletzt zugegriffen am 02.09.2025].

Rességuier, Anaïs und Rowena Rodrigues (2020): „AI ethics should not remain toothless! A call to bring back the teeth of ethics", *Big Data & Society*, Bd. 7(2), S. 1–5, DOI: 10.1177/2053951720942541.

Ribeiro, Marco Tulio; Sammer Singh und Carlos Guestrin (2016): „'Why should I trust you?' Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, S. 1135–1144, https://doi.org/10.1145/2939672.2939778.

Ribeiro, Barbara; Lars Bengtsson, Paul Benneworth, Susanne Bührer, Elena Castro-Martínez, Meiken Hansen, Katharina Jarmai, Ralf Lindner, Julia Olmos-Peñuela, Cordula Ott und Philip Shapira (2018): „Introducing the dilemma of societal alignment for inclusive and responsible research and innovation", *Journal of Responsible Innovation*, Bd. 5(3), S. 316–331, DOI: 10.1080/23299460.2018.1495033.

Rich, Elaine A. (1983): „Artificial Intelligence", McGraw-Hill.

Robbins, Scott (2019): „A Misdirected Principle with a Catch: Explicability for AI", *Minds and Machines*, Bd. 29, S. 495–514, https://doi.org/10.1007/s11023-019-09509-3.

Robert Koch-Institut (2020): „Gesundheitliche Lage der Frauen in Deutschland", Gesundheitsberichterstattung des Bundes gemeinsam getragen von RKI und Destatis, https://www.rki.de/DE/Themen/Gesundheit-und-Gesellschaft/Gesundheitsberichterstattung/Berichte/Frauenbericht/Gesundheitliche_Lage_der_Frauen_2020.pdf?__blob=publicationFile&v=3 [zuletzt zugegriffen am 02.09.2025].

Roeser, Sabine (2014): „The unbearable uncertainty paradox", *Metaphilosophy*, Bd. 45(4–5), S. 640–653, https://doi.org/10.1111/meta.12110.

Rössler, Beate (2001): „Der Wert des Privaten". Frankfurt a. M.: Suhrkamp.

Rössler, Beate (2003): „Anonymität und Privatheit", in: Bäumler, Helmut und Albert von Mutius (Hrsg.): „Anonymität im Internet – Grundlagen, Methoden und Tools zur Realisierung eines Grundrechts", Springer Fachmedien Wiesbaden, S. 27–40.

Rössler, Beate. (2004): „The value of privacy", Polity Press.

Rooksby, Emma (2009): „How to be a responsible slave: managing the use of expert information systems", *Ethics and Information Technology*, Bd. 11, S. 81–90, DOI 10.1007/s10676–009–9183–0.

Ropohl, Günter (2016): „Verantwortung und Risiko", in: Heidbrink, Ludger; Claus Langbehn und Janina Sombetzki (Hrsg.): „Handbuch Verantwortung", Springer Reference Sozialwissenschaften, S. 1–22, DOI 10.1007/978–3–658–06175–3_44–1.

Rudin, Cynthia (2019): „Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence*, Bd. 1, S. 206–215, https://doi.org/10.1038/s42256-019-0048-x.

Rule, James B. (2019): „Contextual Integrity and its Discontents: A Critique of Helen Nissenbaum's Normative Arguments", *Policy & Internet*, Bd. 33(3), S. 260–279, https://doi.org/10.1002/poi3.215.

Russel, Stuart und Peter Norvig (2021): „Artificial Intelligence – A Modern Approach", 4. Auflage, Global Edition, Pearson.

Russo, Frederica; Eric Schliesser und Jean Wagemans (2024): „Connecting ethics and epistemology of AI", *AI & Society*, Bd. 39, S. 1585–1603, https://doi.org/10.1007/s00146-022-01617-6.

Ryan, Mark und Bernd Carsten Stahl (2021): „Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications", *Journal of Information, Communication and Ethics in Society*, Bd. 19(1), S. 61–86, DOI 10.1108/JICES-12–2019–0138.

Sandel, Michael J. (2009): „Gerechtigkeit – Wie wir das Richtige tun", Berlin: Ullstein Buchverlage.

Sasi, Tinshu; Arash Habibi Lashkari, Rongxing Lu, Pulei Xiong und Shahrear Iqbal (2024): „A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges", *Journal of Information and Intelligence*, Bd. 2, S. 455–513, https://doi.org/10.1016/j.jiixd.2023.12.001.

Schmitz, Markus (2020): „Auf dem Weg zu einer ethisch verantwortungsvollen Datennutzung", *Informatik Spektrum*, Bd. 43, S. 366–373, DOI: 10.1007/s00287-020-01298-9.

Searle, John R. (1980): "Minds, brains, and programs", *Behavioral and Brain Sciences*, Bd. 3(3), S. 417–424, DOI: https://doi.org/10.1017/S0140525X00005756.

Shah, Syed Iftikhar Hussain; Vasilis Peristeras und Ioannis Magnisalis (2020): „Government (Big) Data Ecosystem: Definition, Classification of Actors, and Their Roles", *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, Bd. 14(4), S. 102–114, waset.org/Publication/10011154.

Shapiro, Stewart (1997): „Philosophy of Mathematics – Structure and Ontology", Oxford University Press.

Shin, Dong-Hee (2016): „Demystifying big data: Anatomy of big data developmental process", *Telecommunications Policy*, Bd. 40, S. 837–854, http://dx.doi.org/10.1016/j.telpol.2015.03.007.

Sidgwick, Henry (2011) [1874]: „The Methods of Ethics", Cambridge University Press.

Silva, Luciana Maines da; Claudia Cristina Bitencourt, Kadígia Faccin und Tatiana Iakovleva (2019): „The role of stakeholders in the context of responsible innovation: A meta-synthesis." *Sustainability*, Bd. 11(6), S. 1–25, doi:10.3390/su11061766.

Smith, Angela (2005): „Responsibility for attitudes: Activity and passivity in mental life", *Ethics*, Bd. 115(2), S. 236–271, https://doi.org/10.1086/426957.

Snoek Henkemans, A. Franscisca und Jean H.M. Wagemans (2012): „The Reasonableness of Argumentation from Expert Opinion in Medical Discussions: Institutional Safeguards for the Quality of Shared Decision Making", in: Goodwin, Jean (Hrsg.): „Between scientists & citizens: Proceedings of a conference at Iowa State University", Great Plains Society for the Study of Argumentation (GPSSA), S. 345–354.

Solove, Daniel J. (2006): „A taxonomy of privacy", University of Pennsylvania Law Review, Bd. 154(3), S. 477–560, https://ssrn.com/abstract=667622.

Sparrow, Robert (2007): „Killer Robots", *Journal of Applied Philosophy*, Bd. 24(1), S. 62–77, https://doi.org/10.1111/j.1468-5930.2007.00346.x.

Spiekermann, Sarah (2012): „The Challenges of Privacy by Design", *Communications of the ACM*, Bd. 55(7), S. 38–40, https://doi.org/10.1145/2209249.2209263.

Stahl, Bernd Carsten; Simisola Akintoye, Lise Bitsch, Berit Bringedal, Damian Eke, Michele Farisco, Karin Grasenick, Manuel Guerrero, William Knight, Tonii Leach, Sven Nyholm, George Ogoh, Achim Rosemann, Arleen Salles, Julia Trattnig & Inga Ulnicane (2021): „From Responsible Research and Innovation to responsibility by design", *Journal of Responsible Innovation*, Bd. 8(2), S. 175–198, DOI: 10.1080/23299460.2021.1955613.

Stahl, Bernd Carsten; Doris Schroeder and Rowena Rodrigues (2023): „Ethics of Artificial Intelligence – Case Studies and Options for Addressing Ethical Challenges", Springer.

Steffen, Nico; Lukas Wiewiorra und Peter Kroon (2021): „Wettbewerb und Regulierung in der Plattform- und Datenökonomie", *WIK Diskussionsbeitrag*, No. 481, WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste, Bad Honnef.

Stein, Edith (2014): „Freiheit und Gnade – und weitere Texte zu Phänomenologie und Ontologie (1917 bis 1937)", Herder.

Steup, Matthias and Ram Neta (2020): „Epistemology", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (Hrsg.), https://plato.stanford.edu/entries/epistemology/ [zuletzt zugegriffen am 11.10.2023].

Stilgoe, Jack; Richard Owen und Phil Macnaghten (2013): „Developing a framework for responsible innovation", *Research Policy*, Bd. 42, S. 1568–1580, https://doi.org/10.1016/j.respol.2013.05.008.

Strawson, Peter (1962): „Freedom and Resentment." *Proceedings of the British Academy*, Bd. 48, S. 1–25.

Streibich, Karl-Heinz (2015): „Softwareindustrie im Umbruch: Das digitale Unternehmen der Zukunft", in: Linnhoff-Popien, Claudia; Michael Zaddach und Andreas Grahl (Hrsg.): „Marktplätze im Umbruch – Digitale Strategien für Services im Mobilen Internet", Springer Vieweg, S. 15–18.

Strubell, Emma; Ananya Ganesh und Andrew McCallum (2019): „Energy and policy considerations for deep learning in NLP", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, S. 3645–3650, https://doi.org/10.48550/arXiv.1906.02243.

Sweney, Mark (2023): „Royal Mail resumes overseas deliveries via post offices after cyber-attack", The Gusrdian, https://www.theguardian.com/business/2023/feb/21/royal-mail-international-deliveries-cyber-attack-ransom-strikes [zuletzt zugegriffen am 30.04.2025].

Thomson, Judith Jarvis (1975): „The Right to Privacy", *Philosophy & Public Affairs*, Bd. 4(4), S. 295–314, https://www.jstor.org/stable/2265075.

Traub, Jonas; Zoi Kaoudi, Jorge-Arnulfo Quiané-Ruiz und Volker Markl (2020): „Agora: Bringing Together Datasets, Algorithms, Models and More in a Unified Ecosystem", *SIGMOD Record*, Bd. 49(4), S. 6–11, https://doi.org/10.1145/3456859.3456861.

Traum, Anne; Chrostoph Müller, Henning Hummert und Friedemann W. Nerdinger (2017): „Digitalisierung – Die Perspektive des arbeitenden Individuums", White Paper Series, Nr. 1, BMBF- und ESF-gefördertes Verbundprojekt KODIMA, Universität Rostock, https://rosdok.uni-rostock.de/file/rosdok_document_0000010748/rosdok_derivate_0000094718/Traum_Digitalisierung_2017.pdf [zuletzt zugegriffen am 02.09.2025].

Tutt, Andrew (2017): „An FDA for algorithms", *Administrative Law Review*, Bd 69(1), S. 83–123, http://dx.doi.org/10.2139/ssrn.2747994.

Tzafestas, Spyros G. (2018): „Ethics and Law in the Internet of Things World", *Smart Cities*, Bd. 1, S. 98–120, doi:10.3390/smartcities1010006.

Urquhart, Lachlan und Peter J. Craigon (2021): „The Moral-IT Deck: a tool for ethics by design", *Journal of Responsible Innovation*, Bd. 8(1), S. 94–126, DOI: 10.1080/23299460.2021.1880112.

Vallor, Shannon (2016): „Technology and the Virtues – A Philosophical Guide to a Future Worth Wanting", Oxford University Press.

Van den Hoven, Jeroen (1998): „Moral responsibility, public office and information technology", in: Snellen, Ignatius und Wim van de Donk (Hrsg.): „Public administration in an information age: A handbook", S. 97–112, IOS Press.

Van den Hoven, Jeroen (2013): „Value Sensitive Design and Responsible Innovation", in: Owen, Richard; John Bessant and Maggy Heintz (Hrsg.): „Responsible Innovation – Managing the responsible emergence of science and innovation in society", Wiley, S. 27–50 https://doi.org/10.1002/9781118551424.ch4.

Van de Poel, Ibo; Lambèr Royakkers, and Sjoerd D. Zwart (2015): „Moral Responsibility and the Problem of Many Hands", Routledge Studies in Ethics and Moral Theory.

Van Gorp, Anke und Ibo van de Poel (2001): „Ethical Considerations in Engineering Design Processes", *IEEE Technology and Society Magazine*, Bd. 20(3), S. 15–22, DOI: 10.1109/44.952761.

Van Schalkwyk, François; Michelle Willmers und Maurice McNaughton (2016): „Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem", *Information Technology for Development*, Bd. 22(1), S. 68–83, DOI: 10.1080/02681102.2015.1081868.

Van Wynsberghe, Aimee (2021): „Sustainable AI: AI *for* sustainability and the sustainability *of* AI", *AI and Ethics*, Bd. 1, S. 213–218, https://doi.org/10.1007/s43681-021-00043-6.

Vredenburgh, Kate (2022): „The right to explanation", *Journal of Political Philosophy*, Bd. 30(2), S. 209–229, https://doi.org/10.1111/jopp.12262.

Wachter, Sandra und Brent Mittelstadt (2019): „A right to reasonable inferences: re-thinking data protection law in the age of big data and AI", *Columbia Bus Law Review*, Bd. 2, S. 494–620, https://ssrn.com/abstract=3248829.

Warren, Samuel D. und Louis D. Brandeis (1890): „The Right to Privacy", *Harvard Law Review*, Bd. 4(5), S. 193–220.

Watson, Gary (2004): „Agency and Answerability: Selected Essays", Oxford University Press.

Weber, Max (1926): „Politik als Beruf", 2. Auflage, Duncker & Humblot.

Whittlestone, Jess; Rune Nyrup, Anna Alexandrova und Stephen Cave (2019): „The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions", *AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*, January 27–28, 2019, Honolulu, HI, USA, S. 195–200, https: //doi.org/10.1145/3306618.3314289.

Wimmer, Barbara (2019): „Was der neue AMS-Algorithmus für Frauen wirklich bedeutet", future_zone, https://futurezone.at/netzpolitik/was-der-neue-ams-algorithmus-fuer-frauen-wirklich-bedeutet/400617302 [zuletzt zugegriffen am 05.03.2025].

Winikoff, Michael (2018): „Towards Trusting Autonomous Systems", in: El Fallah-Seghrouchni, Amal; Alessandro Ricci und Tran Cao Son (Hrsg.): „Engineering Multi-Agent Systems", *Proceedings of 5th International Workshop, EMAS 2017*, Lecture Notes in Artificial Intelligence, Bd. 10738, S. 3–20, Springer, DOI:10.1007/978-3-319-91899-0_1.

Winkler, Till und Sarah Spiekermann (2018): „Twenty years of value sensitive design: a review of methodological practices in VSD projects", *Ethics and Information Technology*, Bd. 23, S. 17–21, https://doi.org/10.1007/s10676-018-9476-2.

Weiner, Matthew (2005): „Must we know what we say?", *Philosophical Review*, Bd. 114(2), S. 227–251, https://www.jstor.org/stable/30043667.

Westin, Alan Furman (1967): „Privacy and freedom", New York: Atheneum.

Wolf, Susan (2001): „The moral of moral luck", *Philosophical Exchange*, Bd. 31(1), S. 4–19, http://hdl.handle.net/20.500.12648/3203.

Wolff, Jonathan (2021): „Risk and the Regulation of New Technologies", in: Matsuda Tsuyoshi; Jonathan Wolff und Takashi Yanagawa (Hrsg.): „Risks and Regulation of New Technologies", Kobe University Monograph Series in Social Science Research, Springer, S. 3–18, https://doi.org/10.1007/978-981-15-8689-7_1.

Wolff, Jonathan und Avner De-Shalit (2007): „Disadvantage", Oxford University Press, https://doi.org/10.1093/acprof:oso/9780199278268.001.0001.

Wood, Allen W. (1999): „Kant's Ethical Thought", Cambridge University Press.

Young, Iris Marion (2011): „A Social Connection Model", in: Young, Iris Marion und Martha Nussbaum (contributor): „Responsibility for Justice", Oxford University Press, S. 95–122, https://doi.org/10.1093/acprof:oso/9780195392388.003.0004.

Yu, Harlan und David G. Robinson (2012): „The New Ambiguity of ‚Open Government'", *UCLA Law Review Disourse*, Bd. 59, S. 178–208, http://dx.doi.org/10.2139/ssrn.2012489.

Zainab, Syeda Sana e und Tahar Kechadi (2019): „Sensitive and Private Data Analysis: A Systematic Review", *Proceedings of the 3rd International Conference on Future Networks and Distributed Systems* (*ICFNDS '19*), Association for Computing Machinery, S. 1–11, https://doi.org/10.1145/3341325.3342002.

Zeleti, Fatemeh und Adegboyega Ojo (2014): „Capability Matrix for Open Data", *15th Working Conference on Virtual Enterprises*, PRO-VE 2014: Collaborative Systems for Smart Networked Environments, IFIP Advances in Information and Communication Technology, Springer, Bd. 434, S. 498–509, doi.org/10.1007/978–3–662–44745–1_50.

Zeng, Yi; Enmeng Lu und Cunqing Huangfu (2019): „Linking artificial intelligence principles", *Proceedings of the AAAI Workshop on Artificial Intelligence Safety (AAAI-Safe AI 2019)*, S. 1–4, arXiv preprint, https://arxiv.org/abs/1812.04814.

Zick, Andreas (2020): „Sozialpsychologische Diskriminierungsforschung", in: Scherr, Albert, Aladin El-Mafaalani und Anna Cornelia Reinhardt (Hrsg.): „Handbuch Diskriminierung", Springer Reference Sozialwissenschaften (SRS), S. 1–22, DOI: 10.1007/978–3–658–10976–9_4.

Zubcoff, Jose Jacobo; Llorenc Vaquer, Jose-Norberto Mazón, Francisco Maciá, Irene Garrigós, Andrés Fuster und Jose Vincente Carcel (2016): „The University as an Open Data Ecosystem", *International Journal of Design & Nature and Ecodynamics,* Bd. 11(3), S. 250–257, DOI: 10.2495/DNE-V11-N3–250–257.

Zuber, Niina; Jan Gogoll, Severin Kacianka, Alexander Pretschner und Julian Nida-Rümelin (2022): „Empowered and embedded: ethics and agile processes", *Humanities & Social Science Communications*, Bd. 9(191), S. 1–13, https://doi.org/10.1057/s41599-022-01206-4.

Zuiderwijk, Anneke; Marijn Janssen und Chris Davis (2014): „Innovation with open data: Essential elements of open data ecosystems", *Information Polity*, Bd. 19, S. 17–33, DOI 10.3233/IP-140329.

Zuiderwijk, Anneke; Marijn Janssen, Geerten van de Kaa und Kostas Poulis (2016): „The wicked problem of commercial value creation in open data ecosystems: policy guidelines for governments", *Information Polity*, Bd. 21(3), S. 223–236.

Zweig, Katharina A. (2019): „Algorithmische Entscheidungen: Transparenz und Kontrolle", *Analysen & Argumente*, Digitale Gesellschaft, Nr. 338, Konrad-Adenauer Stiftung, https://www.kas.de/de/analysen-und-argumente/detail/-/content/algorithmische-entscheidungen-transparenz-und-kontrolle [zuletzt zugegriffen am 02.09.2025].