

---

## Nachvollziehbarkeit und Kontrolle algorithmischer Entscheidungen und Systeme

Autoren:  
Serpil Taş  
Lukas Wiewiorra

## Danksagung

Die Autoren bedanken sich bei Prof. Dr. Anna Schneider (Hochschule Fresenius) für die Zusammenarbeit und Unterstützung bei der Durchführung der Studie. Außerdem danken die Autoren Carlotta Fischer, Viktoria Jeschall, Christopher Matz und Christian Offer für ihre Anregungen und Unterstützung bei der Strukturierung und Durchführung der Interviews sowie deren Transkription und Auswertung.

## Impressum

WIK Wissenschaftliches Institut für  
Infrastruktur und Kommunikationsdienste GmbH  
Rhöndorfer Str. 68  
53604 Bad Honnef  
Deutschland  
Tel.: +49 2224 9225-0  
Fax: +49 2224 9225-63  
E-Mail: [info@wik.org](mailto:info@wik.org)  
[www.wik.org](http://www.wik.org)

### Vertretungs- und zeichnungsberechtigte Personen

Geschäftsführerin und Direktorin	Dr. Cara Schwarz-Schilling
Direktor	Alex Kalevi Dieke
Direktor Abteilungsleiter Netze und Kosten	Dr. Thomas Plückebaum
Direktor Abteilungsleiter Regulierung und Wettbewerb	Dr. Bernd Sörries
Leiter der Verwaltung	Karl-Hubert Strüver
Vorsitzender des Aufsichtsrates	Dr. Thomas Solbach
Handelsregister	Amtsgericht Siegburg, HRB 7225
Steuer-Nr.	222/5751/0722
Umsatzsteueridentifikations-Nr.	DE 123 383 795

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Algorithmische Entscheidungssysteme</b>	<b>2</b>
<b>3</b>	<b>Internetnutzende und ihre Wahrnehmung algorithmischer Entscheidungssysteme</b>	<b>9</b>
3.1	Bewusstsein und Wissen über algorithmische Entscheidungen	9
3.2	Einstellungen zu algorithmischen Entscheidungen	10
3.3	Transparenz und Kontrolle algorithmischer Entscheidungen	12
<b>4</b>	<b>Forschungsfragen und Methodik</b>	<b>13</b>
<b>5</b>	<b>Forschungsergebnisse</b>	<b>15</b>
5.1	Bewusstsein für die Existenz und Auswirkungen von Algorithmen	15
5.2	Einstellungen zu und Umgang mit Datenschutz und Personalisierung	17
5.3	Verhaltensänderungen	19
<b>6</b>	<b>Schlussfolgerung</b>	<b>22</b>
	<b>Literaturverzeichnis</b>	<b>23</b>



## 1 Einleitung

„Even Google engineers are confused about Google’s privacy settings“ (Lopatto, 2020, The Verge). Diese und ähnliche Schlagzeilen berichten davon, dass die Benutzeroberfläche der Datenschutzeinstellungen von Google so gestaltet ist, dass Einstellungen geändert werden können, aber komplex genug ist, dass die Verbraucher, und selbst Google-Mitarbeiter, nicht immer nachvollziehen können, wie sie die gewünschten Änderungen vornehmen können.

Internetkonzerne wie Alphabet/Google, Meta, Amazon und Apple sammeln in der Regel eine Vielzahl an Daten, die dazu genutzt werden, den Verbrauchern mit Hilfe von Algorithmen personalisiert Inhalte und Werbung anzuzeigen. Allein diese Tatsache ist nicht jedem Verbraucher bewusst. Frühere Studien zeigen, dass ein geringes Bewusstsein über die Existenz von Algorithmen herrscht. Ihre Funktionsweise ist für noch viel weniger Verbraucher nachvollziehbar. Sie können selten einschätzen, welche Daten über sie erhoben und wie diese verarbeitet werden. Die Anpassung der Datenschutzeinstellungen von Diensten gehört zu den Möglichkeiten, wie Verbraucher Kontrolle über die Sammlung und Verarbeitung von Daten ausüben können.

Ziel dieser Studie ist es, aufbauend auf der vorhandenen Literatur die Kompetenz vor allem junger Menschen in Deutschland, welche in der Regel nicht nur die größten Konsumenten digitaler Technologien und Medien sind, sondern auch mit diesen aufgewachsen sind, in Bezug auf Algorithmen zu untersuchen. Dazu wurden semi-strukturierte Interviews mit Verbrauchern im Alter von 20 bis 27 Jahren durchgeführt, welche zu der Generation der „True Digital Natives“ gehören.

Diese Studie gliedert sich in fünf Abschnitte: Zunächst folgt ein kurzer Überblick über algorithmische Entscheidungssysteme und eine Übersicht bestehender Forschung zum Bewusstsein der Verbraucher über die Existenz algorithmischer Systeme und zu ihren Kenntnissen über deren Wirkungsweise. Anschließend werden die Forschungsmethodik und die Ergebnisse der Untersuchung vorgestellt. Die Studie schließt mit einer Schlussfolgerung.

## 2 Algorithmische Entscheidungssysteme

Algorithmen sind vereinfacht ausgedrückt mathematische Funktionen, die Daten als Input verwenden, um eine Lösung für ein bestimmtes Problem abzuleiten oder ein vorgegebenes Resultat zu erreichen. Sie finden zunehmend Einsatz bei der Entscheidungsfindung und bei Entscheidungsverfahren in vielen sozialen und wirtschaftlichen Bereichen (Cormen et al., 2009; Diakopoulos, 2014; Castelluccia und Le Métayer, 2019; Cowgill und Tucker, 2019). So werden sie unter anderem eingesetzt, um Entscheidungen über die Vergabe von Krediten, die Eignung für einen Arbeitsplatz, die Vergabe von Studienplätzen, die Gewährung von Bewährungsstrafen, Präventivmaßnahmen im Bereich des Kinderschutzes oder bei der Rechtsprechung zu treffen (Cofone, 2019; Cowgill und Tucker, 2019; Orwat, 2019). Auch im Internet sind Algorithmen nicht wegzudenken. Sie bestimmen die Preisgestaltung von Produkten, die Platzierung von personalisierter Online-Werbung und Empfehlungen sowie die Anzeige von Nachrichten sowie anderer Online-Medieninhalte und helfen damit die Informationsflut zu begrenzen und die Nutzenden bei der Suche nach relevanten Inhalten zu unterstützen (Powers, 2017; Swart, 2021).

Algorithmische Entscheidungssysteme können danach unterschieden werden, wie stark der Mensch an den endgültigen Entscheidungen beteiligt ist. *Teilautomatische algorithmische Entscheidungssysteme* dienen beispielsweise vorrangig dazu, Menschen bei ihrer Entscheidung zu unterstützen. Die Entscheidung selber wird von einem Menschen getroffen. Darüber hinaus können algorithmische Entscheidungssysteme auch eingesetzt werden, um *automatisierte Entscheidungen* zu treffen, ohne dass Menschen an der endgültigen Entscheidung beteiligt sind. Diese Art von Entscheidungssystemen wird häufig im Internet angewandt. In beiden Fällen liegt die Nutzung von Algorithmen darin begründet, dass sie Aufgaben schneller und effizienter erledigen als Menschen dies tun könnten. Sie ermöglichen die Erfassung, Verarbeitung und Analyse gewaltiger Daten- und Informationsmengen in kürzester Zeit, welche die kognitiven Fähigkeiten eines Menschen übersteigen würden. Zudem dienen sie dazu, menschliche Fehler zu reduzieren und Verzerrungen bei der Entscheidungsfindung zu verhindern (Cowgill und Tucker, 2019; Mehrabi et al., 2021; Renier et al., 2021). Als Konstrukt aus Formeln und Gleichungen werden Algorithmen als gerechtere und neutrale Entscheidungsträger präsentiert. Menschen hingegen, selbst unter den besten Voraussetzungen, können Fehler machen und weisen eine Vielzahl von impliziten Vorurteilen auf, die in ihre Entscheidungen einfließen können (Cofone, 2019).

Studien zeigen beispielsweise auch, dass Menschen dazu neigen, Urteile oder Empfehlungen von Algorithmen zu bevorzugen, als besser zu bewerten und als weniger verzerrt anzusehen als entsprechende Urteile oder Empfehlungen von Menschen (Logg et al., 2019; Araujo et al., 2020; Bonezzi und Ostinelli, 2021). In 2008 fasst Sundar (2008) dieses Verhalten unter dem Konzept der *Maschinenheuristik* zusammen, das besagt, dass ein Urteil als umso objektiver und frei von (ideologischen) Vorurteilen betrachtet wird, je weniger es anthropomorphisiert ist (Sundar, 2008; Araujo et al., 2020). Dennoch sind Algorithmen nicht frei von Verzerrungen und können gleichfalls, wie die Entscheidungen von Menschen, zu fehlerbehafteten oder gar diskriminierenden Ergebnissen kommen. Zahlreiche Studien und Artikel bestätigen dies.

Einige bekannte Beispiele für verzerrte und diskriminierende Ergebnisse algorithmischer Entscheidungsfindungssysteme beziehen sich auf die ethnische Herkunft. Angwin et al. (2016) finden beispielsweise, dass COMPAS<sup>1</sup> – ein Algorithmus, welcher in vielen US-Bundesstaaten eingesetzt wird, um die Rückfallwahrscheinlichkeit von Straftätern und damit ihre Eignung für eine Freilassung auf Kautions bis zum Beginn des Gerichtsverfahrens zu beurteilen – die Rückfallwahrscheinlichkeit von schwarzen Straftätern systematisch überschätzt. Die Wahrscheinlichkeit, dass schwarze Straftäter fälschlicherweise in

---

<sup>1</sup> COMPAS ist das Akronym für Correctional Offender Management Profiling for Alternative Sanctions.

die Gruppe mit hohem Rückfallrisiko eingestuft werden, obwohl sie danach tatsächlich nicht erneut straffällig werden, ist doppelt so hoch wie bei weißen Straftätern. Lum und Isaac (2016) liefern bei der Untersuchung des "PredPol"-Algorithmus ebenfalls Belege für diskriminierende Ergebnisse. Der Algorithmus wird ebenfalls in den USA verwendet und ermittelt, in welchen Gebieten mit einer hohen Kriminalitätsrate zu rechnen ist, um entsprechend mit Polizeipräsenz zu reagieren. Lum und Isaac (2016) haben den Algorithmus auf die Drogenkriminalität in Oakland in den USA angewendet und festgestellt, dass Stadtteile mit überwiegend afroamerikanischer Bevölkerung oder anderen Ethnien eher als Brennpunkte krimineller Aktivitäten identifiziert werden als Gebiete mit überwiegend weißen Anwohnern, obwohl der tatsächliche Drogenkonsum zwischen den Gruppen im Allgemeinen kaum Unterschiede aufweist. Die Konsequenz eines solchen Ergebnisses ist, dass durch den „PredPol“-Algorithmus vor allem nicht-weiße Gemeinden unverhältnismäßig stark durch die Polizei überwacht werden. Dies könnte bei einem vergleichbaren Drogenkonsum langfristig zu einer überproportionalen Aufdeckung von Drogendelikten in diesen Gebieten führen und damit die durch den Algorithmus verursachte Verzerrung durch neue Daten verstärken.

Auch einige Erkennungssysteme liefern in Bezug auf ethnische Merkmale verzerrte Ergebnisse. Buolamwini und Gebru (2018) führen Experimente mit kommerziellen Gesichtserkennungssystemen durch und zeigen, dass die Gesichter von Menschen mit einem helleren Hautton grundsätzlich besser von diesem System erkannt werden als die von Menschen mit einem dunkleren Hautton. Snow (2018) berichtet von einem Test, bei dem die Gesichtserkennungssoftware „Amazon Rekognition“ verwendet wurde, um Bilder von Kongressabgeordneten der USA mit einer Datenbank von Fahndungsfotos abzugleichen. Nicht-weiße Mitglieder wurden hier überproportional häufig als Straftäter identifiziert. Ähnliches gilt bei der Texterkennung. Blodgett und O'Connor (2017) wenden vier verschiedene Spracherkennungssysteme von namhaften Herstellern auf Nachrichten an, die im sozialen Netzwerk Twitter gepostet wurden. Die Autoren finden Unterschiede in der Genauigkeit bei der Texterkennung, wobei die Genauigkeit bei Textnachrichten, die von weiblichen oder afroamerikanischen Autoren verfasst wurden, schlechter ausfällt als bei weißen männlichen Autoren. Die Unterschiede sind besonders groß bei kurzen Textnachrichten. Da Daten aus sozialen Netzwerken häufig für Meinungsanalysen zu Produkten oder politischen Persönlichkeiten verwendet werden, ist die Tatsache, dass die Meinungen von Afroamerikanern weniger gut erfasst werden als die von weißen Individuen, als ein gesellschaftliches Problem einzustufen. Letztlich sind die automatisierten Entscheidungssysteme, die im Internet Anwendung finden, nicht zwangsläufig frei von ethnischer Diskriminierung. So findet Sweeney (2013) Indizien dafür, dass Online-Anzeigen mit Straftatbezug häufiger bei Suchanfragen mit Namen auftauchen, die stereotypisch schwarz klingen als bei Namen, die stereotypisch weiß klingen. Da viele Menschen das Internet nutzen, um mehr über bestimmte Personen zu erfahren, kann eine Anzeige unter dem Namen der betreffenden Person oder neben ihren Leistungen, die vermuten lässt, dass sie möglicherweise vorbestraft ist, schädlich sein, unabhängig davon, ob dies der Wahrheit entspricht.

Häufig kommt es bei algorithmischen Systemen und deren Entscheidungen auch zu Verzerrungen zwischen den Geschlechtern. Dies wurde von verschiedenen Forschern untersucht und dokumentiert, insbesondere im Online-Kontext. Prates et al. (2020) untersuchen beispielsweise die Verzerrungen bei der maschinellen Übersetzung von Texten, insbesondere im Hinblick auf die geschlechterspezifische Zuordnung von Berufsgruppen. Die Autoren konstruieren Sätze wie "Er/Sie ist ein Ingenieur" in verschiedenen geschlechtsneutralen Sprachen und übersetzen diese mit Hilfe der Google Translate API ins Englische. Sie finden, dass Google Translate dazu neigt, mehrheitlich männliche Pronomen zu verwenden, insbesondere in Bereichen, die typischerweise mit einer unausgewogenen Geschlechterverteilung oder Stereotypen in Verbindung gebracht werden, wie z. B. bei MINT-Berufen (Wissenschaft, Technologie, Ingenieurwesen und Mathematik). Die Autoren vergleichen die Ergebnisse auch mit realen Statistiken über den Anteil von Frauen in den einzelnen Berufen und zeigen, dass die von Google Translate

gelieferten Ergebnisse nicht die tatsächliche Verteilung der weiblichen Beschäftigten widerspiegeln. Insgesamt liefert Google Translate viel häufiger männliche Pronomen, als aufgrund statistischer Daten zu erwarten wäre. Kay et al. (2015) analysieren hingegen die Ergebnisse der Bildersuche zu Berufen, die über die Google-Suchmaschine abgerufen werden. Die Untersuchungsergebnisse weisen unter anderem darauf hin, dass Frauen in den Suchergebnissen für Berufsgruppen, die stereotypisch von Männern dominiert werden, unterrepräsentiert sind, während das Verhältnis der Geschlechter in den offiziellen Beschäftigungsstatistiken für diese Berufsgruppen etwa ausgeglichen ist.

Darüber hinaus zeigen Datta et al. (2015) und Lambrecht und Tucker (2019), dass Online-Stellenanzeigen nicht geschlechterneutral angezeigt werden. Datta et al. (2015) verwenden in ihrer Studie beispielsweise "AdFisher", um die Wechselwirkung zwischen dem Browsing-Verhalten von Nutzenden, den Werbeanzeigen von Google und den Anzeigeneinstellungen zu untersuchen. Sie zeigen unter anderem, dass, wenn das Geschlecht in den "Anzeigeneinstellungen" auf Frauen umgestellt wird, Anzeigen für besser bezahlte Jobs vergleichsweise mehr Männern als Frauen angezeigt werden, und zwar obwohl das Surfverhalten der Nutzenden unverändert bleibt. Lambrecht und Tucker (2019) weisen in einer empirischen Fallstudie ebenfalls geschlechtsspezifische Unterschiede in der Sichtbarkeit von Online-Werbung für MINT-Berufe nach. Hierzu haben die Autoren eine eigene Werbeanzeige kreiert, die in 191 Ländern auf Facebook geschaltet wurde. Insgesamt wurde die Testwerbung 20 Prozent häufiger Männern als Frauen gezeigt, obwohl Frauen mit größerer Wahrscheinlichkeit auf die Werbung reagierten.

Auch die von Unternehmen intern verwendeten Algorithmen scheinen hinsichtlich des Geschlechts bei der endgültigen Einstellungsentscheidung zu diskriminieren. Im Jahr 2018 wurde in den Medien über einen Algorithmus von Amazon berichtet, mit dem das Unternehmen Bewerbende im Hinblick auf ihre Eignung einstuft. Das System zeigte eine klare Voreingenommenheit gegenüber weiblichen Bewerbenden und eine Bevorzugung männlicher Bewerbender. Da der Algorithmus mit den Daten von Amazon zu vergangenen Einstellungen trainiert wurde und davon ausging, dass die aktuelle Mitiberauswahl von Amazon ein guter Indikator für die zukünftige Mitiberauswahl von Amazon ist, spiegelte der Algorithmus die bestehenden Einstellungspraktiken wider (Cofone, 2019). Angwin et al. (2017a) decken darüber hinaus auch Altersdiskriminierung bei Stellenanzeigen auf Facebook auf und dokumentieren, dass Stellenanzeigen von Unternehmen wie Amazon, Verizon, UPS, Goldman Sachs und Facebook älteren Facebook-Nutzenden nicht angezeigt wurden.

Neben diesen Beispielen gibt es zahlreiche weitere Artikel und Studien, die über Verzerrungen und Diskriminierung bei algorithmischen Entscheidungen berichten – sei es auf Webseiten für Freiberufler, wo der Ranking-Algorithmus Frauen und Schwarze deutlich schlechter bewertet als andere Personen mit denselben Qualifikationen oder Angeboten (Hannák et al., 2016; Hannák et al., 2017); in Bezug auf Online-Wohnungsanzeigen, die bestimmten Bevölkerungsgruppen wie Müttern mit Kindern im Teenageralter, Einwanderern oder Menschen jüdischen Glaubens nicht angezeigt werden (Angwin et al., 2017b); in Systemen, die die Allokation von Patenten übernehmen und ethnische Diskriminierung aufweisen (Obermeyer et al., 2019); im Internet, wo Unternehmen Preisdiskriminierung auf Basis von Standortdaten betreiben (Valentino-DeVries et al., 2012) und bei der öffentlichen Verwaltung, die schlicht fehlerhafte Algorithmen zu nutzen scheint (Egan und Roberts, 2021).

Alle diese Beispiele zeigen deutlich, welche Risiken Algorithmen und algorithmische Entscheidungen und Ergebnisse im Zusammenhang mit dem Gleichheitsgrundsatz bergen. Sie zeigen, dass sich das Versprechen eines fairen, rationalen und unvoreingenommenen algorithmischen Entscheiders nicht immer erfüllt. Dies liegt insbesondere daran, dass die Verzerrungen und Diskriminierungen in den Entscheidungen von Algorithmen nach wie vor auf den Menschen zurückzuführen sind. Selbst völlig autonome Algorithmen sind Werkzeuge, die von Menschen gestaltet wurden (Cofone, 2019). Jeder Algorithmus erfordert einen Menschen, der das zu lösende Problem formuliert und eine Entscheidung



darüber trifft, was der Algorithmus vorhersagen soll, sowie welche Trainingsdaten verwendet werden. Die Voreingenommenheit der Menschen, die den Algorithmus programmieren und anwenden, kann in den Algorithmus einfließen (Cofone, 2019; Kleinberg et al., 2019). Selbst bei einem theoretisch unvoreingenommenen Programmierer kann es vorkommen, dass er aufgrund kognitiver Einschränkungen das vorliegende Problem nicht vollständig erfassen kann und sich nur auf eine begrenzte Teilmenge von Variablen konzentriert, die ihm bekannt und vertraut sind, wodurch die Algorithmen fehlerhaft kalibriert werden (Schwartzstein, 2014; Bordalo et al., 2016; Cowgill und Tucker, 2019). Blodgett und O'Connor (2017), welche wie oben beschrieben Unterschiede in der Genauigkeit der Erfassung von Textnachrichten zwischen weiblichen sowie afroamerikanischen Autoren und weißen, männlichen Autoren durch verschiedene Spracherkennungssysteme feststellen, heben hervor, dass Schwarze und Frauen einen geringeren Anteil an den Ingenieurteams, die die Algorithmen für die getesteten Spracherkennungssysteme entwickelten, ausmachten. Wäre der Pool an Programmierern diverser gewesen, so die Vermutung, wären diese Probleme erkannt worden. Cowgill et al. (2020) stellen dies sogar empirisch fest. Dazu werten sie 8,2 Millionen Ergebnisse von Algorithmen von etwa 400 Programmierern aus. Obwohl die Autoren Verzerrungen in den Ergebnissen hauptsächlich auf Verzerrungen in den Trainingsdaten zurückführen, stellen sie auch fest, dass die Verzerrungen mit der Demografie der Programmierer korrelieren, insbesondere mit dem Geschlecht. Sie finden beispielsweise, dass es wahrscheinlicher ist, dass die Verzerrungen in den Ergebnissen männlicher Programmierer miteinander korrelieren. Besteht ein Programmierer-Team aus mehr männlichen Programmierern, verstärkt sich der jeweilige Verzerrungsfehler. Sie schlussfolgern daher, dass diversere Programmier-Teams diesem Effekt entgegen wirken können.

Erfüllt der Algorithmus zudem einen monetären Zweck, wird er nicht nur direkt durch den Programmierer beeinflusst, sondern auch durch Vorgaben, denen die betreibenden Unternehmen unterliegen. Auf dieses Problem weisen unter anderem Ali et al. (2019) hin. Die Autoren haben untersucht, ob Algorithmen zu einer verzerrten Auslieferung von Online-Werbung führen – d. h., dass einige Nutzende aufgrund ihrer demografischen Merkmale weniger wahrscheinlich Werbung sehen als andere –, die nicht auf die Targeting-Entscheidungen der Werbetreibenden zurückgeführt werden kann. Zu diesem Zweck haben die Autoren mehrere Werbekampagnen und Anzeigen auf Facebook geschaltet. Die Auswertung zeigt, dass eine Auslieferung auf Facebook unter anderem von Markt- und finanziellen Optimierungseffekten sowie der plattformeigenen Vorhersage über die "Relevanz" von Anzeigen für verschiedene Nutzergruppen abhängt. Sie stellen fest, dass sowohl das Budget des Werbetreibenden als auch der Inhalt der Anzeige erheblich zur verzerrten Auslieferung von Facebook-Anzeigen beiträgt. Kritisch anzumerken ist, dass sie trotz neutraler Targeting-Parameter eine signifikante Verzerrung bei der Auslieferung von "echten" Anzeigen für Arbeits- und Wohnungsangebote entlang von Geschlecht und Ethnie beobachten. Die Ergebnisse zeigen Mechanismen auf, die zu einer potenziell diskriminierenden Anzeigenauslieferung führen können, selbst wenn Werbetreibende ihre Targeting-Parameter so einstellen, dass sie möglichst umfassend sind. Zu einem ähnlichen Schluss kommen auch Lambrecht und Tucker (2019), die experimentell eine verzerrte Auslieferung von Stellenanzeigen für MINT-Berufe feststellen. Die Autoren interpretieren das Ergebnis dahingehend, dass das Anzeigen von Werbung für die Gruppe "Frauen" teurer ist. Das bedeutet, dass ein Werbealgorithmus, der die Werbeschaltung auf Kosteneffizienz optimiert, Werbung, die eigentlich geschlechtsneutral sein sollte, auf diskriminierende Weise ausliefern kann.

Abbildung 2-1: Quellen von Verzerrungen im Algorithmus



Quelle: Eigene Darstellung; angelehnt an: Leslie et al. (2021), S.373. Icons von Gulherme Furtade (links oben), IconPai (rechts oben) Eko Purnomo (rechts unten), Flatart (links unten) (Noun Project)

Eine weitere Ursache für verzerrte Ergebnisse bei algorithmischen Entscheidungen ist die Auswahl von Trainingsdaten. Wenn ein Algorithmus verzerrte oder unvollständige Datensätze verwendet bzw. mit diesen trainiert wird, wird ein Algorithmus auch verzerrte Ergebnisse produzieren. Aber auch wenn der Trainingsdatensatz vollständig, unverzerrt oder gar repräsentativ ist, besteht die Möglichkeit, dass ein Algorithmus verzerrte Ergebnisse erzeugt. Diese entstehen, wenn in den Daten soziale Ungleichheiten und vergangene Diskriminierung bereits eingebettet sind (Cofone, 2019; Kleinberg et al., 2019). Wenn ein Algorithmus auf Basis historischer Datensätze operiert wie beispielsweise COMPAS oder PredPol in der Strafverfolgung, bei denen schwarze Personen oder andere Minderheiten aufgrund von eingebetteten gesellschaftlichen Ungleichheiten im Durchschnitt mehr Verhaftungen aufweisen als beispielsweise weiße Personen, wird der Algorithmus auch zukünftig diskriminierende Ergebnisse liefern. Die darauf aufbauenden menschlichen Handlungen (z.B. verstärkte Kontrollen) können diesen Effekt dann auch zukünftig durch beispielsweise verzerrte Kriminalstatistiken zementieren. Ähnliches gilt bei Einstellungsentscheidungen. Nutzt ein Algorithmus eine Datenbank bzw. wird ein Algorithmus auf Basis einer Datenbank mit überwiegend männlichen Mitarbeitenden trainiert, wie es beispielsweise bei Amazon der Fall war, hat der Algorithmus gelernt, besser über die zu erwartende Produktivität männlicher Bewerbender zu entscheiden als über die weiblicher Bewerbender. Wenn die Daten also repräsentativ für die Realität sind, werden sie auch bestehende Verzerrungen widerspiegeln. Ein schwerwiegendes Risiko bei der algorithmischen Entscheidungsfindung sind vor allem Verzerrungen, denen nicht entgegengewirkt wird und die daher Gefahr laufen, systemimmanent zu werden, da die Resultate bereits verzerrter Entscheidungen wieder in den Algorithmus eingespeist werden (Cofone, 2019).

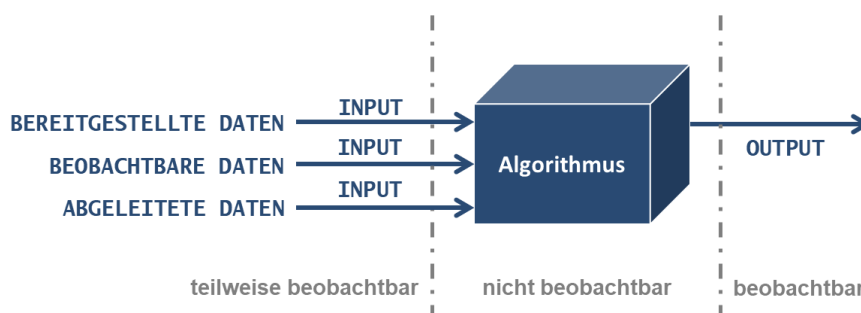
Aus diesen Gründen ist die Anwendung von Algorithmen in einigen Bereichen stark umstritten. Zivilgesellschaftliche Akteure fordern die EU-Gremien beispielsweise auf, die Verwendung von prädiktiven und profilierenden KI-Systemen bzw. Algorithmen in der Strafverfolgung zu verbieten (Krempf, 2022). Diese Arten von KI-Systemen bzw. Algorithmen würden nach dem aktuellen Entwurf des Gesetzes über künstliche Intelligenz der Europäischen Kommission<sup>2</sup> als hochriskant eingestuft. Gemäß Artikel 6 Absatz 2 des Gesetzesentwurfs würden KI-Systeme, die bei der biometrischen Identifizierung und Kategorisierung natürlicher Personen, der Verwaltung und dem Betrieb kritischer Infrastrukturen, der allgemeinen und beruflichen Bildung, der Beschäftigung, dem Personalmanagement und dem Zugang zur

<sup>2</sup> Europäische Kommission (2021). Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final, 2021/0106(COD).

Selbstständigkeit eingesetzt werden, als *KI-Systeme mit hohem Risiko* eingestuft. Algorithmen oder KI-Systeme, die bei Entscheidungen über den Zugang zu und die Inanspruchnahme von grundlegenden privaten und öffentlichen Diensten und Leistungen eingesetzt werden, wie z. B. die Berechtigung zur Inanspruchnahme von Unterstützungsleistungen, Krediten oder Nothilfe, werden im aktuellen Entwurf ebenfalls als hochriskant eingestuft. Das Gleiche gilt für Entscheidungen im Zusammenhang mit Migration oder Asyl und dem Einsatz von KI-Systemen bei der Grenzkontrolle. Schließlich sind auch KI-Systeme, die Justizbehörden bei der Ermittlung und Auslegung von Tatsachen und Rechtsnormen sowie bei der Anwendung des Rechts auf bestimmte Sachverhalte unterstützen, laut dem aktuellen Entwurf hochriskant. Viele der eingangs diskutierten problematischen Algorithmen wären also in die Hochrisikokategorie einzuordnen. Der Gesetzentwurf sieht strenge Auflagen für Anbieter von KI-Systemen mit hohem Risiko vor. Diese müssten nach Artikel 10 und 17 ihre KI-Systeme mit Trainings-, Validierungs- und Testdaten füttern, die bestimmte Qualitätsanforderungen erfüllen und durch geeignete Data-Governance- und Managementverfahren verwaltet werden. Ferner unterlägen sie Auflagen nach Artikel 15 zur Gewährleistung eines angemessenen Maßes an Genauigkeit, Robustheit und Cybersicherheit ihrer Systeme und müssten nach Artikel 9 ein robustes Risikomanagement einrichten und anwenden.

Unabhängig davon ist ein weiteres wesentliches Problem, dass Algorithmen von außen betrachtet eine Black-Box sind. Die genaue Wirkungsweise entzieht sich der Beobachtbarkeit. Während sowohl der Input an Daten als auch der Output zumindest zu einem gewissen Grad beobachtet werden können, ist die Funktionsweise intransparent. Diese Undurchsichtigkeit resultiert zum Teil aus den Bemühungen der Unternehmen, ihre Systeme zu schützen und geheim zu halten, um zu verhindern, dass Dritte sie ausnutzen, sowie aus den begrenzten technischen Kenntnissen der Nutzenden, um Algorithmen zu erfassen und zu verstehen (Burrell, 2016). Selbst die Entwickler von Algorithmen wissen oft nur teilweise, wie ihre Algorithmen insgesamt funktionieren. Das liegt daran, dass es sich bei den meisten Algorithmen um Kombinationen von Codestücken handelt, die von mehreren Entwicklern gemeinsam konzipiert wurden und bei denen eine große Anzahl von Berechnungsverfahren und Variablen zusammenwirken. Dadurch ist die zugrunde liegende Syntax selbst für hochqualifizierte Personen schwer zu entschlüsseln (Swart, 2021).

Abbildung 2-2: Das Black-Box-Problem



Quelle: Eigene Darstellung.

Abhilfe könnte hier vor allem mehr Transparenz schaffen, welche sowohl der Entwurf zum Gesetz über künstliche Intelligenz als auch der Entwurf der Richtlinie über KI-Haftung der Europäischen Kommission<sup>3</sup> zukünftig vorsieht. So sollen nach dem Gesetzentwurf zur künstlichen Intelligenz Anbieter von risikoreichen KI-Systemen künftig eine detaillierte technische Dokumentation erstellen müssen, die

<sup>3</sup> Europäische Kommission (2022). Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates zur Anpassung der Vorschriften über außervertragliche zivilrechtliche Haftung an künstliche Intelligenz (Richtlinie über KI-Haftung), COM(2022) 496 final, 2022/0303 (COD).

Informationen über den Verwendungszweck, die einzelnen Komponenten der KI, ihren Entwicklungsprozess, die Überwachung, den Betrieb, die Kontrolle, das Risikomanagementsystem und weitere Angaben zu dem KI-System enthält. KI-Systeme mit hohem Risiko müssten auch die automatische Aufzeichnung von Vorgängen und Ereignissen während des Betriebs ermöglichen, um die Rückverfolgbarkeit der Ergebnisse zu gewährleisten und potenzielle Risiken zu verfolgen. Sämtliche Informationen und Unterlagen, die erforderlich sind, um die Konformität der Hochrisiko-KI-Systeme mit den Anforderungen der Richtlinie nachzuweisen, wären auf Verlangen der zuständigen nationalen Behörde zu übermitteln. Auch gegenüber den Nutzenden und anderen natürlichen Personen müssten die Anbieter der Hochrisiko-KI-Systeme nach dem Entwurf zum Gesetz über künstliche Intelligenz Transparenzanforderungen erfüllen. Demnach müssten Hochrisiko-KI-Systeme so konzipiert und entwickelt sein, dass ihr Betrieb für die Nutzenden hinreichend transparent ist. In entsprechenden Gebrauchsanweisungen wären Informationen zum Anbieter und zu den Merkmalen, Fähigkeiten und Leistungsgrenzen des Hochrisiko-KI-Systems erforderlich. Letztlich müssten die Anbieter – von einigen Ausnahmen abgesehen – sicherstellen, dass KI-Systeme, die für die Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass natürlichen Personen signalisiert wird, dass sie mit einem KI-System interagieren.

Der Entwurf der Richtlinie über KI-Haftung befasst sich im Wesentlichen mit Haftungsfragen für Schäden, die durch den Einsatz eines KI-Systems entstehen. Die Richtlinie ermöglicht Gerichten den Zugang zu Informationen über bestimmte Hochrisiko-KI-Systeme, bei denen der Verdacht besteht, dass sie einen Schaden verursacht haben (z.B. zur Feststellung von Schadensersatzansprüchen).

Mit dem Gesetz über digitale Dienste der Europäischen Kommission (nachfolgend: Digital Services Act, DSA)<sup>4</sup> ist zudem bereits eine Verordnung in Kraft, die zumindest für Algorithmen und KI-Systeme, die von Plattformbetreibern entwickelt und angewandt werden, mehr Transparenz bringen soll. Insbesondere die Anbieter sehr großer Online-Plattformen sind nach dem Gesetz über digitale Dienste verpflichtet, in ihren Allgemeinen Geschäftsbedingungen in klarer und verständlicher Sprache die wichtigsten Parameter, die in ihren Empfehlungsalgorithmen verwendet werden, und alle Möglichkeiten für die Nutzenden, diese zu ändern, darzulegen. In Bezug auf Online-Werbung wird vorgeschrieben, dass die Anbieter Informationen z. B. über den Inhalt der Werbung, den Werbetreibenden, den Werbezeitraum und die Parameter, die verwendet werden, um den Nutzenden der Plattform die Werbung zu liefern, zusammenstellen und über Anwendungsprogrammierschnittstellen für den gesamten Zeitraum, in dem sie eine Werbung anzeigen, und für ein Jahr nach der letzten Anzeige der Werbung über ihre Systeme öffentlich zugänglich machen. Darüber hinaus unterliegen Anbieter sehr großer Online-Plattformen gegenüber Behörden verschiedenen Informations- und Rechenschaftspflichten. So müssen sie auf begründetes Verlangen dem Koordinator für digitale Dienste und zugelassenen Forschenden Zugriff auf Daten gewähren, die für die Überwachung und Bewertung der Einhaltung dieser Verordnung erforderlich sind bzw. die zur Aufspürung, zur Ermittlung und zum Verständnis systemischer Risiken beitragen. Sie sind verpflichtet, dem nationalen Koordinator für digitale Dienste Berichte über Risikobewertungen, nötige Abhilfemaßnahmen bei Eintreten von Risiken und externe Prüfungen zu übermitteln. Inzwischen richtet die Kommission auch ein Europäisches Zentrum für Algorithmische Transparenz (ECAT) ein, „um ihre Aufsichtsfunktion mit internem und externem multidisziplinärem Wissen zu unterstützen“ (Europäische Kommission, 2022). Das Zentrum soll die Kommission „bei der Bewertung unterstützen, ob die Funktionsweise algorithmischer Systeme mit den Risikomanagementverpflichtungen übereinstimmt, die der DSA [...] festlegt“ (Europäische Kommission, 2022).

---

<sup>4</sup> Europäische Kommission (2022). Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19.10.2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), Amtsblatt der Europäischen Union, L 277/1.

### 3 Internetsnutzende und ihre Wahrnehmung algorithmischer Entscheidungssysteme

Auf Basis der im vorherigen Kapitel vorgestellten Grundlagen stellt sich die Frage, wie Individuen algorithmische Entscheidungen als Empfänger bzw. Betroffene wahrnehmen. Wissen sie von der verbreiteten Anwendung von Algorithmen? Können sie ihre Funktionsweisen begreifen? Nehmen sie Verzerrungen in algorithmischen Entscheidungen wahr? Haben sie das Gefühl, Kontrolle über Algorithmen zu haben?

Es gibt eine wachsende Zahl wissenschaftlicher Arbeiten, die sich mit Algorithmen und den aufgeführten Fragen beschäftigen. Die verschiedenen Forschungsstränge befassen sich unter anderem mit der Untersuchung des Wissens über die Funktionsweise und des Bewusstseins über die Existenz algorithmischer Entscheidungen sowie mit der Akzeptanz solcher Entscheidungen. Dabei gelangen die Studien zu teils gegensätzlichen Ergebnissen.

#### 3.1 Bewusstsein und Wissen über algorithmische Entscheidungen

Gegensätzliche Studienergebnisse finden sich insbesondere hinsichtlich des Bewusstseins über die Existenz und des Wissens über die Funktionsweise von Algorithmen bei Individuen. Grzymek und Puntschuh (2019) führen zum Beispiel eine allgemeine, europaweite Umfrage zu diesem Thema durch. Sie schlussfolgern, dass die Hälfte der europäischen Bevölkerung weder weiß, was ein Algorithmus ist, noch, dass sie bereits in vielen Lebensbereichen eingesetzt werden. Die andere Hälfte hat nur ein rudimentäres Wissen über Algorithmen. In einer früheren Studie mit Facebook-Nutzenden stellen Eslami et al. (2015) ebenfalls fest, dass sich mehr als die Hälfte der Teilnehmenden der Existenz des Kuratierungsalgorithmus im Facebook News Feed überhaupt nicht bewusst ist. Die Teilnehmenden führen das Fehlen von Benachrichtigungen in ihren Facebook News Feeds häufig entweder auf die Entscheidung ihrer Freunde und Familien, sie auszuschließen, oder auf ihre eigenen Aktivitäten zurück, aber nicht auf einen Algorithmus. Die Studie von Rader und Gray (2015), die im selben Jahr ebenfalls das Verständnis der Nutzenden für die algorithmische Kuratierung im News-Feed von Facebook näher untersucht, kommt zu einem ganz anderen Ergebnis. Rader und Gray (2015) finden heraus, dass die meisten Facebook-Nutzenden sich tatsächlich bewusst sind, dass sie nicht alle Beiträge ihrer Freunde sehen. Ein Großteil der Teilnehmenden gibt an, zu glauben, dass eine Instanz, die als Facebook oder als Algorithmus bezeichnet wird, Beiträge für die Anzeige im News-Feed priorisiert. Viele verknüpfen sogar die angezeigten Inhalte mit dem, was das System anscheinend über das Verhalten, die Präferenzen, die früheren Interaktionen und die Eigenschaften der Nutzenden sowie die Beliebtheit der Beiträge weiß. Bucher (2017) befasst sich ebenfalls mit dem Wissen von Individuen über Algorithmen im Zusammenhang mit dem News-Feed von Facebook. In der Studie wird gezeigt, dass obwohl die Teilnehmenden der Studie ausdrücklich angeben, nicht zu wissen, was ein Algorithmus ist, sie dennoch mehr oder weniger ausgefeilte Theorien darüber haben, was der Facebook-Algorithmus tut. Einige vermuten auch hier, dass er als Reaktion auf alle von Facebook gesammelten Daten funktioniert, und verstehen intuitiv, dass der Facebook-Algorithmus eine Verbindung zwischen ihren Online-Aktivitäten und den Arten von Beiträgen herstellt, die ihnen angezeigt werden. Einige haben zudem eine klare Vorstellung davon, wie sie sich verhalten müssen, um den Algorithmus zu lenken. Dies zeigt unter anderem, dass die Unkenntnis einer bestimmten Terminologie oder ein Mangel an technologischem Vokabular nicht zwangsläufig dazu führt, dass es einer Person an algorithmischem Bewusstsein mangelt (Swart, 2021). Siles et al. (2019) finden zum Beispiel heraus, dass Netflix-Nutzende das Wort *Algorithmus* nur selten verwenden, aber Hypothesen darüber aufstellen, warum der Dienst bestimmte Empfehlungen anzeigt. Auch die Internetsnutzenden in der Studie von Dogruel et al. (2020) haben, unabhängig von ihrer Vertrautheit mit dem Begriff Algorithmus, zumindest eine vage Vorstellung von der Funktionsweise von Algorithmen im

Internet. Die Teilnehmenden erkennen Algorithmen vor allem in Situationen, in denen sie mit personalisierten Inhalten konfrontiert werden, die sie eindeutig auf ihr eigenes Online-Verhalten zurückführen können. Ein Ergebnis, welches sich auch in Studien von Grzymek und Puntschuh (2019) wiederfindet. Das meiste Wissen über den Einsatz von Algorithmen stellen die Autoren ebenfalls im Bereich der Personalisierung von Werbung fest. Allerdings scheint zu der Anwendung von Algorithmen in anderen Bereichen, insbesondere zu Anwendungen außerhalb der Online-Welt, wenig Bewusstsein zu herrschen. Zu einem anderen Ergebnis kommt hingegen Bucher (2017). In einer Studie wird festgestellt, dass das algorithmische Bewusstsein durch Überraschung oder Bestürzung ausgelöst wird. Dies kann zum Beispiel eintreffen, wenn der Algorithmus falsche Klassifizierungen vornimmt oder fehlerhafte Vorhersagen macht und damit unzutreffende Inhalte ausspielt.

Die Wahrnehmung von und die Kenntnisse über Algorithmen scheinen von demografischen Merkmalen beeinflusst zu sein. Sowohl Grzymek und Puntschuh (2019), Dogruel et al. (2020), Cotter und Reisdorf (2020) als auch Gran et al. (2021) finden hierzu Unterschiede in Bezug auf Alter, Bildung und Geschlecht. Doch auch diese Ergebnisse sind nicht eindeutig. So schlussfolgern Grzymek und Puntschuh (2019), dass im Wesentlichen Männer, ältere Individuen und Individuen mit einem hohen Bildungsniveau besser über Algorithmen informiert sind. In den Arbeiten von Dogruel et al. (2020) und Cotter und Reisdorf (2020) sind es vor allem jüngere Personen, die Algorithmen verstehen. Gran et al. (2021) kommen zum selben Ergebnis. In ihrer Studie stellen sie fest, dass der Bekanntheitsgrad von Algorithmen in jüngeren Altersgruppen am höchsten ist. Die Ergebnisse in Bezug auf Geschlecht und Bildung sind jedoch ähnlich wie die von Grzymek und Puntschuh (2019).

### 3.2 Einstellungen zu algorithmischen Entscheidungen

Isoliert betrachtet, haben Individuen als Empfänger algorithmischer Entscheidungen eine zwiespältige Haltung gegenüber Algorithmen. Sie entwickeln offenbar sowohl positive als auch negative Assoziationen mit ihnen. Studien zeigen unter anderem, dass die Einstellung des Einzelnen zu algorithmischen Entscheidungen von individuellen Merkmalen wie Bildungsstand, Alter, Geschlecht, von der persönlichen Überzeugung in Bezug auf Datenschutz und Privatsphäre, von der Wahrnehmung und dem Verständnis von Algorithmen im Allgemeinen und letztlich vom jeweiligen Kontext abhängt, in dem sie algorithmischen Entscheidungen begegnen.

Ein Zusammenhang zwischen individuellen Merkmalen der Probanden und der Wahrnehmung der Fairness und Nützlichkeit automatisierter Entscheidungs- und Empfehlungssysteme wird beispielsweise von Araujo et al. (2020) aufgezeigt. So stellen die Autoren fest, dass ein hohes Maß an Allgemeinbildung sowie spezielles Wissen über Algorithmen sich positiv auf die Wahrnehmung der Nützlichkeit und Fairness automatisierter Entscheidungs- und Empfehlungssysteme auswirkt. Datenschutzbedenken im Allgemeinen haben einen negativen Einfluss auf die Wahrnehmung der Nützlichkeit und Fairness automatisierter Entscheidungs- und Empfehlungssysteme und veranlassen Menschen dazu, die Systeme als riskant einzustufen. Je größer die Datenschutzbedenken, desto negativer ist die Einstellung gegenüber den algorithmischen Entscheidungssystemen. Je stärker die Teilnehmenden jedoch an ihre eigenen Fähigkeiten glauben, ihre Privatsphäre zu schützen, desto positiver ist ihre Wahrnehmung der Nützlichkeit und Fairness von algorithmischen Entscheidungssystemen und desto geringer ist das wahrgenommene Risiko. Die demografischen Merkmale Alter und Geschlecht haben bei Araujo et al. (2020) ebenfalls einen Einfluss auf die Wahrnehmung von Algorithmen. Auch das Gefühl der Autonomie kann bei algorithmischen Entscheidungen unterschiedlich wahrgenommen werden. Dogruel et al. (2020) finden in ihrer Studie zwei Gruppen: Individuen, die das Gefühl haben, dass Algorithmen ihre Autonomie fördern, und Individuen, die das Gefühl haben, in ihrer Autonomie eingeschränkt zu werden. Teilnehmende der Studie, die in die erstere Kategorie fallen, gehen bei ihren Beschreibungen auf Situationen ein, in

denen sie beispielsweise Suchmaschinen verwendet haben. Algorithmen werden hier als hilfreiche Werkzeuge wahrgenommen, die eine Informationsüberflutung verhindern und den Nutzenden helfen, durch eine Vorauswahl relevante Inhalte auszuwählen; ein Phänomen, das auch in anderen Studien thematisiert wurde. So stellte Christin (2017) fest, dass die Reduktion der Komplexität zu einer effizienteren Informationsverarbeitung führt, wovon die Nutzenden profitieren. Die andere Gruppe von Teilnehmenden äußert jedoch Bedenken hinsichtlich der Auswirkungen von Algorithmen auf ihre Entscheidungsfindung und empfindet sie als manipulativ. Dieses Gefühl tritt häufig auf, wenn die Teilnehmenden über ihre Erfahrungen beim Online-Shopping berichten. Dort werden Algorithmen als Einschränkung wahrgenommen, insbesondere in Bezug auf die Produktauswahl, die für sie in Betracht gezogen wird. Swart (2021) isoliert ebenfalls verschiedene Gruppen von Individuen, die sich in ihrer Einstellung zu Algorithmen unterscheiden. Teilnehmende der Studie, die Algorithmen als neutralen, mathematischen und rationalen Prozess verstehen, zeigen in der Regel keine starke emotionale Reaktion auf Algorithmen. Für diese Gruppe sind sie lediglich Mittel zum Zweck. Die zweite Gruppe betrachtet Algorithmen als nützliche Wegweiser, die relevante Informationen hervorheben. Diese Teilnehmenden betonten ähnlich wie bei Dogruel et al. (2020) vor allem die Vorteile von Empfehlungssystemen wie Zeitersparnis und Effizienz. Bei der dritten Gruppe, die Swart (2021) identifiziert, rufen Algorithmen stark negative Gefühle hervor. Diese betrachtet Algorithmen als mächtige Impulsgeber, die sie insbesondere zu Käufen anregen möchten und dahingehend manipulieren.

Inwieweit werden Entscheidungen von Algorithmen im Vergleich zu menschlichen Entscheidungen anders wahrgenommen? In der Literatur über *algorithmic aversion* finden Forscher häufig Verhaltensweisen und Einstellungen, die auf ein grundlegendes Misstrauen gegenüber algorithmischen Entscheidungen hindeuten. Yeomans et al. (2019) vergleichen die Akzeptanz von Empfehlungen, die von Menschen gemacht werden, mit denen von Algorithmen. Sie finden unter anderem, dass automatisierte Empfehlungssysteme im Vergleich zu menschlichen Entscheidungsträgern als undurchschaubar empfunden werden können. Dies kann sich auf die Bereitschaft von Personen auswirken, das System oder seine Empfehlungen zu akzeptieren. Die Teilnehmenden der Studie von Dietvorst et al. (2015) neigen dazu, Algorithmen abzulehnen und sich für menschliche Entscheidungsträger zu entscheiden, trotz der überlegenen Leistung von Algorithmen. Außerdem sind Individuen gegenüber automatisierten Entscheidungsfindungssystemen weit weniger nachsichtig als gegenüber Menschen, wenn beide Entscheidungsträger denselben Fehler machen bzw. ähnlich ungenaue Ergebnisse liefern (Dietvorst et al., 2015; Bogert et al., 2021; Renier et al., 2021; Langer et al., 2022). Darüber hinaus stellen Bucher (2017) und Grzymek und Puntschuh (2019) fest, dass sich einige Menschen bei rein computerbasierten Entscheidungen grundsätzlich unwohl fühlen und von Menschen getroffene Entscheidungen vorziehen.

In der Literatur über *algorithm appreciation* werden Algorithmen hingegen tendenziell als gerechtere und zuverlässigere Entscheidungsträger wahrgenommen. Das Vertrauen rührt insbesondere daher, dass die von Algorithmen getroffenen Entscheidungen auf objektiven Daten und der Verwendung von Statistiken beruhen und daher besser sind als subjektive Meinungen (Dodge et al., 2019). So sind die Teilnehmenden in der Studie von Bigman et al. (2022) weniger empört über algorithmische Diskriminierung als über menschliche Diskriminierung. Insgesamt führen die Autoren sechs Untersuchungen durch, um die Empörung über verschiedene Arten der algorithmischen Diskriminierung (Geschlecht, Alter, ethnische Zugehörigkeit) bei Jobeinstellungen zu testen. Die Asymmetrie in der Empörung ist insbesondere darauf zurückzuführen, dass Maschinen weniger voreingenommene Motive zugeschrieben werden. Daraus ergibt sich, dass Menschen eine algorithmische Diskriminierung eher akzeptieren als eine menschliche Diskriminierung. Ein ähnliches Ergebnis leiten Bonezzi und Ostinelli (2021) ab. Sie befassen sich mit algorithmischen Entscheidungen bei Hochschulzulassungen, der Gewährung von Bewährungen und bei Einstellungen. Sie stellen fest, dass algorithmische Entscheidungen, die zu geschlechterbedingten oder ethnischen Ungleichheiten führen, weniger wahrscheinlich als verzerrt wahrgenommen werden als

menschliche Entscheidungen. Dies liegt daran, dass Menschen glauben, dass Algorithmen, im Gegensatz zu Menschen, die Entscheidungsfindung dekontextualisieren, indem sie individuelle Merkmale vernachlässigen und Regeln und Verfahren blind anwenden, unabhängig davon, über wen sie urteilen. Diese Überzeugung führt dazu, dass die Menschen glauben, dass Algorithmen eher als Menschen alle Menschen gleich behandeln und daher weniger wahrscheinlich voreingenommene Entscheidungen treffen. Folglich können Situationen, in denen Diskriminierung vorkommt, unkritisch akzeptiert werden.

### 3.3 Transparenz und Kontrolle algorithmischer Entscheidungen

Algorithmen sind in der Regel undurchsichtig – vor allem für Individuen, die die Empfänger bzw. Betroffenen algorithmischer Entscheidungen sind. Transparenz kann Individuen in die Lage versetzen, fundierte Entscheidungen darüber zu treffen, wie sie mit einem Algorithmus bzw. Dienst, der einen Algorithmus verwendet, interagieren und die potenziellen Folgen der getroffenen Entscheidungen beurteilen. Transparenz schafft ebenso ein Gefühl der Kontrolle und die Möglichkeit, das Machtverhältnis zwischen den Betreibern und Anwendern von Algorithmen und den von den jeweiligen Entscheidungen betroffenen Personen in ein Gleichgewicht zu bringen oder zu verlagern (Rader et al., 2018).

Mögliche Transparenzmaßnahmen untersuchen beispielsweise Rader et al. (2018), Farke et al. (2021) und Arias-Cabarcos et al. (2022). Im Rahmen eines Online-Experiments legen Rader et al. (2018) den Teilnehmenden verschiedene Erläuterungen über die Existenz und Funktionsweise des News-Feed-Algorithmus von Facebook vor. Alle Erklärungen machen den Teilnehmenden zwar die Funktionsweise des Systems bewusster und helfen ihnen, die Korrektheit und Genauigkeit der Systeme zu beurteilen, doch wecken sie bei den Teilnehmenden im Allgemeinen nicht das Gefühl, dass sie die Algorithmen in irgendeiner Weise kontrollieren können – weder über die jeweilige Benutzeroberfläche noch durch ihr eigenes Verhalten.

Arias-Cabarcos et al. (2022) und Bemmann et al. (2022) konzentrieren sich in ihren Studien auf die Nutzung von Datenschutz-Dashboards. Datenschutz-Dashboards schaffen Transparenz, indem sie den Nutzenden ermöglichen, die Datenerfassung und den Datenschutz von Online-Diensten zu erkunden und zu verwalten und damit im Prinzip auch die Daten, die von Algorithmen verwendet werden können. Vor allem im Online-Kontext verstehen die Betroffenen häufig nicht, in welchem Umfang ihre eigenen Daten gesammelt, hergeleitet und erzeugt und wie die Daten eingesetzt werden. Arias-Cabarcos et al. (2022) führen eine Studie unter Facebook-Nutzenden durch, bei der der Schwerpunkt auf dem Transparenz-Dashboard "Off-Facebook Activity" liegt, das über die von Dritten gesammelten Daten informiert und deren Verwaltung ermöglicht. Im Allgemeinen sind die Studienteilnehmenden von der großen Menge an gesammelten Daten überrascht und fühlen sich mit der Datenerhebung eindeutig unwohl. Sie geben auch an, dass sie in Zukunft angemessenere Schutzmaßnahmen ergreifen möchten, wie z. B. sich besser über die Einstellungen auf Facebook zu informieren, verstärkt private Einstellungen auf Facebook zu konfigurieren und die Weitergabe ihrer Daten durch Facebook-Partner einzuschränken. Obwohl die Teilnehmenden es gut finden, dass das Dashboard ihnen die Möglichkeit bietet, ihren Verlauf zu löschen, ihre Daten herunterzuladen und künftige Aktivitäten zu verwalten, und sie die Einfachheit der Benutzeroberfläche und ihre leichte Bedienbarkeit und Verständlichkeit loben, missfällt ihnen vor allem der Mangel an Informationen darüber, wie Daten gesammelt werden, von welchem Gerät Daten gesammelt werden und welche Daten genau weitergegeben werden. Darüber hinaus weisen mehrere Nutzende darauf hin, dass es sich um ein Opt-in-Verfahren und nicht um ein Opt-out-Verfahren handeln sollte. Bemängelt wird ebenfalls das mangelnde Bewusstsein über die Existenz des Tools und wie schwierig es zu finden sei.



Farke et al. (2021) untersuchen My Activity von Google in einem ähnlichen Kontext. Die Teilnehmenden in dieser Studie erkennen einen signifikant höheren Nutzen der Datenerfassung durch Google und haben weniger Bedenken, nachdem sie mit My Activity interagiert haben. Insgesamt erhöht die Interaktion bei den Teilnehmenden das Gefühl der subjektiven (und objektiven) Kontrolle über die gesammelten Daten und verringert die Bedenken hinsichtlich des anfänglichen Gefühls der mangelnden Kontrolle und der Unfähigkeit, den Datenzugriff einzuschränken. Negativ ist jedoch, dass My Activity den Nutzenden nicht dabei hilft, fundierte Entscheidungen über einzelne Aktivitäten zu treffen, da die Auswirkungen der Aufbewahrung oder Löschung einzelner Daten für den Nutzenden undurchsichtig sind.

## 4 Forschungsfragen und Methodik

Die vorherigen Abschnitte dieser Studie legen nahe, dass Algorithmen zwar effizienzsteigernd sind, aber nicht immer zu wünschenswerten, verzerrungsfreien Ergebnissen führen. Die Betroffenen von algorithmischen Entscheidungen stehen zumeist im Dunkeln – nicht nur bezüglich der Existenz dieser Systeme, sondern auch in Bezug darauf, welche Daten verarbeitet werden und wie dies geschieht. Im Folgenden werden daher drei Forschungsfragen untersucht, die insgesamt das Bewusstsein und die Wahrnehmung von Verbrauchern in Bezug auf den Einsatz sowie die Funktionsweisen algorithmischer Entscheidungssysteme adressieren:

- Forschungsfrage 1: Wissen Verbraucher von der Existenz algorithmischer Entscheidungssysteme? (Bewusstsein für den Algorithmus)
- Forschungsfrage 2: Wissen Verbraucher, welche Daten von ihnen erfasst und durch Algorithmen verarbeitet werden? (Funktionsweise von Algorithmen)
- Forschungsfrage 3: Ist es den Verbrauchern möglich, die Datenerfassung und -verarbeitung im gewünschten Maße zu kontrollieren? (Kontrolle von Algorithmen)

Diese qualitative, explorative Untersuchung stützt sich auf semi-strukturierte Einzelinterviews. Die Wahl dieses Ansatzes ermöglicht es, die Einstellungen, Erfahrungen und das Verhalten von Individuen in Bezug auf Algorithmen in einem natürlichen Umfeld und aus ihrer eigenen Perspektive zu erfassen und zu beobachten (Patton, 2002; Bryman, 2003; Mohajan, 2018). Ziel ist es, ein detailliertes Verständnis der komplexen Interaktionen zwischen Menschen und Algorithmen zu erlangen.

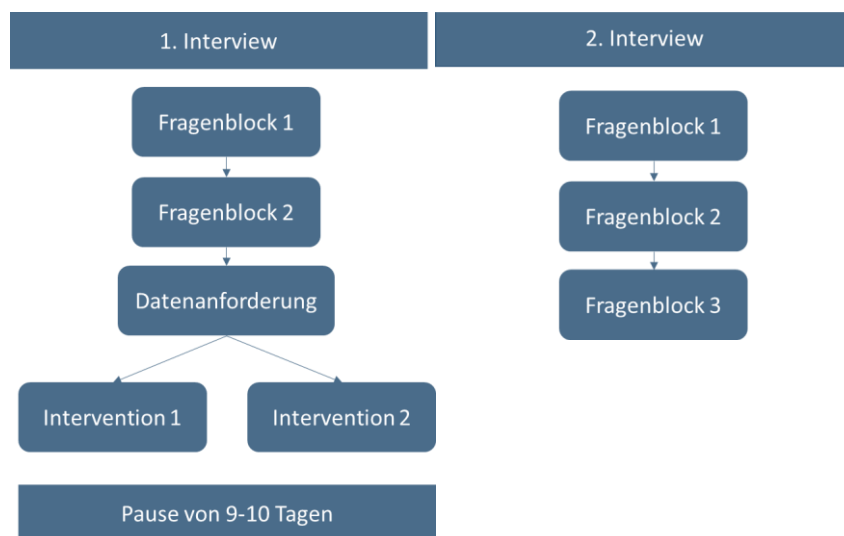
Insgesamt wurden 20 in Deutschland ansässige Verbraucher zwischen 20 und 27 Jahren interviewt. Die Mehrheit der Teilnehmenden waren Studierende (60 %). Es nahmen neun männliche (45 %) und elf weibliche (55 %) Personen an den Interviews teil. Auch wenn eine solche Stichprobe keineswegs repräsentativ für die deutschen Internetnutzer ist, bietet die hier untersuchte Generation Z als "True Digital Natives" eine ideale Möglichkeit, Algorithmen und ihre Wahrnehmung durch eben diese Nutzenden zu untersuchen. Die Generation zeichnet sich insbesondere dadurch aus, dass sie im Vergleich zu vorherigen Generationen in einem technologisch geprägten Umfeld aufgewachsen ist, das durch eine zunehmende digitale, soziale und personalisierte Mediennutzung gekennzeichnet ist (Swart, 2021).

Der Fokus der Untersuchung lag dabei auf dem Instagram-Algorithmus. Hamilton et al. (2014) argumentieren, dass sich das Bewusstsein für Algorithmen am besten messen lässt, indem man den Befragten einen bestimmten Algorithmus präsentiert und dann Fragen stellt, die die Wahrnehmung der algorithmischen Prozesse messen. Instagram gehört zu den bekanntesten sozialen Netzwerken des Unternehmens *Meta*, welches bei jüngeren Menschen besonders beliebt ist. Insbesondere bei der hier betrachteten Altersgruppe liegt der Anteil der Nutzenden bei etwa 80 % (Statista, 2022). Bis auf einen Teilnehmenden, der seine Nutzung auf Instagram mit "mehrmals die Woche" beschreibt, liegt die Nutzungsintensität bei allen anderen 19 Teilnehmenden bei "mehrmals täglich".

Wie viele andere Plattformen sammelt Instagram Daten der Nutzenden. Dazu gehören sowohl Informationen, die der Nutzende im Benutzerkonto bereitstellt, wie der Name, die Telefonnummer, das Alter, das Geschlecht etc., als auch Informationen, die durch die Nutzung des Dienstes und anderer verknüpfter Dienste anfallen. Darüber hinaus werden Geräteinformationen und Cookie-Daten erhoben sowie Informationen, die durch Partner von Instagram, wie Werbetreibende, App-Entwickler und -Publisher, übermittelt werden.<sup>5</sup>

Die Teilnehmenden nahmen im Frühjahr 2022 an je zwei Online-Interviews teil, die über die Plattform Zoom durchgeführt wurden. Die Daten wurden in Kooperation mit der Hochschule Fresenius erhoben. Die Interviews wurden mit Hilfe von Studierenden, die an der Durchführung der Interviews mitwirkten, manuell transkribiert.

Abbildung 4-1: Ablauf der Interviews



Quelle: Eigene Darstellung.

Das erste Interview hatte durchschnittlich eine Länge von 30 Minuten und unterteilte sich in vier Blöcke. Der erste Fragenblock umfasste hauptsächlich Fragen, die herausstellen sollten, ob die Teilnehmenden personalisierte Inhalte wahrnehmen und ob sie einen Zusammenhang zwischen ihrem Verhalten, den von Instagram gesammelten Daten, einem Algorithmus und den angezeigten Inhalten herstellen können.

Der daran anschließende zweite Fragenblock befasste sich hauptsächlich mit dem Thema Datenschutz. Das primäre Ziel war, sowohl die grundsätzliche Einstellung der Teilnehmenden zum Thema Datenschutz zu erfassen als auch Informationen zu ihrer Datenschutzstrategie auf Instagram zu erhalten.

Im Anschluss wurden die Teilnehmenden gebeten, ihre bei Instagram gespeicherten Daten anzufordern. Der Datenexport diente als Grundlage für das zweite Interview. Während der Durchsicht hatten die Teilnehmenden die Möglichkeit, ihre Gedanken zu verbalisieren und frei zu berichten. Im Rahmen der Interviews wurden insbesondere drei Ordner aus dem Datenexport näher betrachtet: die Ordner „Messages“, „Shopping“ und „Ads and Businesses“.

<sup>5</sup> Informationen zu den Daten, die Instagram erhebt, finden sich auf der folgenden Seite: <https://help.instagram.com/155833707900388> [letzter Zugriff: 05.09.2022].

Zuletzt wurden die Teilnehmenden zufällig in zwei Interventionsgruppen aufgeteilt. Den Teilnehmenden einer der beiden Gruppen wurde eine Aufklärung zu Instagram und seinem Algorithmus präsentiert. Diese verfolgte das Ziel, die Teilnehmenden für die Konsequenzen von Algorithmen zu sensibilisieren. Die Teilnehmenden der anderen Gruppe erhielten eine Aufklärung zur Änderung verschiedener Datenschutzeinstellung auf Instagram. Es fand somit eine Befähigung zur Konfiguration der Einstellungen auf Instagram statt. Die Einstellungsmöglichkeiten, die in den Interviews genauer besprochen wurden, waren die Einstellungen für Cookies, Werbeanzeigen, Einstellungen rund um den Feed und Einstellungen, die direkt am Gerät der Teilnehmenden vorgenommen werden können.

Das zweite anschließende Interview, welches in der Regel etwa 9 bis 15 Tage nach dem ersten Interview durchgeführt wurde und im Durchschnitt etwa 15 Minuten dauerte, diente dazu, Veränderungen in der Wahrnehmung von Algorithmen und dem Umgang mit ihnen nach dem ersten Interview und den erhaltenen Erkenntnissen zu erfahren. Gleichsam wurden Reaktionen der Teilnehmenden auf die von Instagram gesammelten Daten erhoben.

Im Folgenden werden die Ergebnisse der inhaltlichen Analyse der Interviews dargestellt. Die Namen aller 20 Teilnehmenden sind pseudonymisiert.

## 5 Forschungsergebnisse

Dieses Kapitel ist entlang der in Kapitel 4 dargestellten Forschungsfragen gegliedert. Zunächst werden die Ergebnisse hinsichtlich des Bewusstseins und des Verständnisses der Teilnehmenden in Bezug auf den Algorithmus und die Praktiken der Datenerfassung von Instagram vorgestellt. Anschließend werden die Einstellungen der Teilnehmenden zum Datenschutz im Allgemeinen und zu den von Instagram bereitgestellten Datenschutzeinstellungen dargestellt. Abschließend wird erörtert, welche Maßnahmen die Teilnehmenden aufgrund ihrer Interaktion mit den Datenschutzeinstellungen und der Datenerfassungspraxis von Instagram in Zukunft zu ergreifen beabsichtigen.

### 5.1 Bewusstsein für die Existenz und Auswirkungen von Algorithmen

Eines der wesentlichen Ziele dieser Arbeit ist es, herauszufinden, ob sich die Verbraucher der Existenz algorithmischer Systeme bewusst sind oder zumindest vermuten, dass es solche Systeme gibt. Im Rahmen der Interviews wird deutlich, dass nahezu alle Teilnehmenden zu wissen scheinen, dass Instagram ein automatisiertes Verfahren verwendet, um ihnen personalisierte Inhalte, Anzeigen und Empfehlungen auszuspielen. Sie nehmen intuitiv eine gewisse Verbindung zwischen ihrem Online-Verhalten auf Instagram und den Inhalten, die ihnen angezeigt werden, wahr. In diesem Zusammenhang vermuten alle Teilnehmenden, dass Instagram Daten über sie sammelt und diese zur Ausspielung der personalisierten Inhalte verwendet. Obwohl einige Kategorien von Daten von den Teilnehmenden während der Interviews korrekt identifiziert werden, benennen sie nur einen Bruchteil der Informationen, die Instagram tatsächlich über sie sammelt. Einzelnen betrachtet haben die Teilnehmenden im Allgemeinen nur ein oberflächliches und unvollständiges Wissen über die gesammelten Daten und erwähnen in der Regel jeweils nur eine oder zwei relevante Kategorien von Daten. Sie gehen davon aus, dass insbesondere Informationen über gelikte Beiträge, die jeweilige Verweildauer auf bestimmten Beiträgen, die Kanäle, denen sie folgen oder die ihnen folgen, Korrespondenzen mit anderen Nutzern und Interaktionen mit Beiträgen verwendet werden, um ihre Interessen abzuleiten und Inhalte entsprechend anzupassen. Nur die Hälfte der Teilnehmenden geht darüber hinaus noch darauf ein, dass auch außerhalb von Instagram Daten über sie anfallen und von Instagram verwendet werden können. So beschreibt eine teilnehmende Person: „Die wissen auch, was ich alles generell im Internet suche [...]. Zum Beispiel: Wenn ich irgendwas google, dann kriege ich direkt fünf Minuten später eine Werbung genau dafür.“ (ITN12). Ähnlich

beschreibt eine andere teilnehmende Person: „*Ich glaube, dass [...], wenn ich irgendwas google [...], Instagram das irgendwie auch mitbekommt. Wenn ich in bestimmten Zeiten sehr viele Rezepte google, [...] kommt es sehr oft vor, dass mir dann auf Instagram sehr viele Kochvideos angezeigt werden*“ (ITN20).

Lediglich acht der 20 Teilnehmenden lassen den Begriff *Algorithmus* in ihren Ausführungen fallen. Die meisten von ihnen mit dem Zusatz, dass sie nicht genau wissen, wie dieser funktioniert. Viele der Erklärungen, die die Teilnehmenden für Algorithmen geben, bestehen aus wenig technischem Vokabular und beruhen auf ihren Erfahrungen, wie zum Beispiel im Fall dieser Person: „*Ich stelle mir das so vor: [...] wenn ich mir [einen Beitrag] lange anschau, dann ist es natürlich so, dass der Algorithmus merkt, dass das für mich interessant ist und mir weiter angezeigt werden kann. Wenn ich einfach nur weiter scrolle, dann merkt [der Algorithmus], dass das für mich irrelevant ist.*“ (ITN02) Ähnlich beschreibt es auch diese teilnehmende Person: „*Der Algorithmus wird anhand dessen angepasst, was ich mir gerade angeschaut habe. [...] Zum Beispiel: Ich schaue mir ein Video zu einem bestimmten Thema an und das fünfte [Video] danach ist schon direkt wieder zu dem Thema und immer so weiter. Also, je mehr ich mir zu einem Thema angucke, desto mehr wird mir auch angezeigt [...]*“ (ITN15). Genau drei Teilnehmende versuchen sich an eher technischen Definitionen. Diese Teilnehmenden beschreiben Algorithmen ganz allgemein als mathematische Formeln, computergestützte mathematische Programme oder Operationen. So sagt eine teilnehmende Person beispielsweise: „*[...] ein Algorithmus versucht deine Präferenzen zu speichern, umzusetzen, nach Inhalten zu suchen, die [...] den Inhalt haben, der dir gefallen könnte [...]. Algorithmus allgemein ist, denke ich, irgendwie so eine wiederkehrende Folge von Zahlen oder irgend sowas in der Computersprache. Damit kenne ich mich jetzt nicht genug aus.*“ (ITN11). Eine andere Person führt noch etwas weiter aus und erklärt den Begriff Algorithmus in Zusammenhang mit den eigenen Kenntnissen zu künstlicher Intelligenz: „*Ein Algorithmus ist irgendwie für mich erstmal, irgendwo im Bereich Mathematik, Computer Science, künstliche Intelligenz [anzusiedeln]. Irgendwie stelle ich mir das so vor, dass [...] man erstmal eine leere Box hat, die mit allem gefüllt wird, was man quasi da reinwirft – wann oder um wieviel Uhr benutze ich die App, was gucke ich mir an, was habe ich schon mal gesucht. [...] Der Algorithmus versucht dann auf jeden Fall erstmal, dass man sich auf der App wohlfühlt, dass man die App mehr nutzt. [...] Ich glaube auch, dass wenn ich vielleicht jedes Jahr einmal nach Fahrrädern suche und vielleicht sogar immer zur gleichen Zeit, das der Algorithmus auch sowas lernt.*“ (ITN19).

Zusammenfassend lässt sich festhalten, dass die Teilnehmenden zwar ein gewisses Verständnis dafür haben, dass die ihnen von Instagram vorgeschlagenen Inhalte eine Antwort auf alle von Instagram gesammelten Daten sind, dass ihnen aber im Allgemeinen das technische Vokabular und Wissen fehlt, um die zugrundeliegende Mechanik genau zu benennen, ähnlich wie in früheren Studien. Die Erklärungen basieren meist auf den Erfahrungen der Befragten während der Nutzung. Auch die Kenntnis über die grundsätzliche Breite und Tiefe der Daten, die Instagram von ihnen sammelt, ist eher oberflächlich – ein Ergebnis, welches sich mit früheren Studien deckt (Rader, 2014; Rao et al., 2014; Herder und van Maaren, 2020). Dies zeigt sich vor allem auch an den Reaktionen der Befragten, wenn sie die von Instagram gesammelten Daten betrachten.

Wie im vorangegangenen Kapitel beschrieben, wurden die Teilnehmenden gebeten, ihre bei Instagram gespeicherten Daten anzufordern. In den meisten Fällen sind die Teilnehmenden erstaunt über die Menge an Daten gewesen, die Instagram ihnen über sie selbst zur Verfügung stellt. Sie haben zwar erwartet, dass Daten über sie gespeichert werden, aber nicht in dem Umfang, der ihnen präsentiert wurde. Insbesondere der Ordner mit der Aufschrift „Messages“, der im Vergleich zu den beiden anderen betrachteten Ordnern bei jedem der Teilnehmenden die meisten Informationen enthält, hat bei den Teilnehmenden unterschiedliche Emotionen hervorgerufen. Die meisten Teilnehmenden sind negativ

überrascht und verwenden in ihren Ausführungen Adjektive wie „erschreckend“, „beängstigend“ oder „beunruhigend“. Eine dieser Personen fasst ihre Gedanken wie folgt zusammen: „Boah!“ [...] „Dieses Boah, ist diese Verblüffung, wie sehr das in meine Privatsphäre eindringt. [...] Ein negatives Gefühl. [...] schon erschreckend [...].“ (ITN03). Auffällig ist, dass die Teilnehmenden verwundert sind, wie weit die Daten zurückgehen. Dieselbe Person führt beispielsweise aus: „Es hat mich überrascht, dass da auf einmal Nachrichten oder Kommentare von 2015 noch von mir abgespeichert sind. [...] Mich hat es einfach umgehauen, dass die das sichern, also es ist ja sieben Jahre her.“ (ITN03). Nur zwei Teilnehmende schienen von der Datenaufbereitung begeistert zu sein und äußern weniger negativ konnotierte Empfindungen.

## 5.2 Einstellungen zu und Umgang mit Datenschutz und Personalisierung

Die meisten Befragten haben eine eher laxen Einstellung zu Datenschutz. Viele der Teilnehmenden geben zu, sich bisher kaum Gedanken bezüglich Datenschutz im Allgemeinen oder im Zusammenhang mit Instagram gemacht zu haben. Häufig zeigt sich, dass die Teilnehmenden für das Thema Datenschutz Gleichgültigkeit empfinden. Vereinzelt wird dieses Gefühl auch mit der Empfindung von Machtlosigkeit kommuniziert. Wenn soziale Dienste wie Instagram verwendet werden, ist die Tatsache, dass Daten über sie anfallen, die wiederum aufgegriffen, verarbeitet und genutzt werden, unumgänglich. Sie sehen in der Regel zwei Optionen, entweder stimmen sie den Datenschutzrichtlinien und Geschäftsbedingungen der Anbieter zu oder entsagen der Nutzung:

*„Ich muss sagen, dass ich mir über Datenschutz eigentlich gar nicht so viele Gedanken mache. Ich weiß, man sollte das eigentlich, aber ich halte das für utopisch [...]. [...] man kann ja gar nicht NICHT mitmachen, dann grenzt man sich selber aus. [...] dann haben sie halt meine Daten.“ (ITN01)*

*„Ich sag mal so, um eine App nutzen zu können, muss man zwingend den [vorgegebenen] Datenschutzbedingungen zustimmen. Das heißt, man hat ja eigentlich eher weniger die Wahl, wenn man diese App wirklich nutzen möchte. [...] Es ist schon wichtig zu wissen, was da drinsteht, aber man kann halt nicht großartig was dran ändern.“ (ITN04)*

Einige begründen ihre Einstellung zum Thema Datenschutz auch damit, dass die Materie zu komplex ist, um sie umfassend zu verstehen. Darstellungen und Informationen zu Datenschutz – angenommen die Teilnehmenden glauben, dass es sie gibt – sind ihrer Meinung nach nicht zugänglich oder verständlich. Der Umgang mit Datenschutzbestimmungen und -einstellungen wird als sehr zeitaufwendig und komplex empfunden. Die Teilnehmenden wollen und / oder können sich in ihrem Alltag nicht damit auseinandersetzen. Zudem scheinen die Teilnehmenden weniger besorgt darüber zu sein, dass die Plattformanbieter ihre Daten erfassen und weiterverwerten, als dass andere Dritte auf ihre Daten zugreifen könnten und die betreffenden Informationen öffentlich preisgeben.

Dies zeigt auch, dass sich die Mehrheit der Teilnehmenden noch nicht näher mit den Datenschutzeinstellungen von Instagram auseinandergesetzt hat. Kaum einer der Teilnehmenden hat Änderungen an den Standardeinstellungen vorgenommen – bis auf die Tatsache, dass sie ihre Accounts auf privat gestellt haben. Dies deckt sich mit der Befürchtung, dass andere Dritte die Daten missbrauchen könnten. Die einzige zusätzliche Einstellungsänderung, die drei Teilnehmende benennen, ist das Ausschalten aller Cookies bis auf die erforderlichen.

Obwohl sich nur sehr wenige Teilnehmende tatsächlich mit den Datenschutz-Einstellungen auseinandergesetzt haben, gehen die meisten Teilnehmenden bereits davon aus, dass sie wahrscheinlich nicht die Möglichkeit haben werden, die Einstellungen nach ihrem persönlichen Bedarf anzupassen. Sie sind

der Meinung, dass es zwar verschiedene Einstellungsmöglichkeiten gibt, diese aber unzureichend sind oder nur den Anschein erwecken sollen, dass Änderungen in den Einstellungen möglich sind. In diesem Zusammenhang sprechen mehrere Teilnehmende auch die Unübersichtlichkeit des Zugangs zu diesen Einstellungen an, da diese oft versteckt und schwer zu finden sind. Einige Teilnehmende sind jedoch zuversichtlicher und gehen davon aus, dass Instagram kein Interesse daran hat, kryptische Informationen oder Einstellungstools anzubieten oder sogar gesetzlich verpflichtet ist, bestimmte Einstellungsmöglichkeiten bereitzustellen.

In der Tat scheint es so zu sein, dass die Einstellungsoptionen für die Datenerfassung oder in Bezug auf algorithmische Entscheidungen den Teilnehmenden nicht zugänglich und verständlich präsentiert werden. Wie in der Methodik beschrieben, wurden einige der Befragten darüber aufgeklärt, wie sie verschiedene Datenschutzeinstellungen auf Instagram ändern können, mussten diese aber erst selbst finden.

Seit dem Frühjahr dieses Jahres erlaubt Instagram seinen Nutzern, ihren Feed wieder chronologisch, ohne den Einfluss eines Algorithmus, anzeigen zu lassen. Keiner der Teilnehmenden hat diese Einstellung bisher genutzt. Darüber hinaus gibt die Mehrheit an, noch nie von der Funktion gehört zu haben. Sie haben nicht das Gefühl, dass es von Instagram Informationen gab, die direkt auf diese Einstellungsmöglichkeit hinwiesen. Entsprechend schwierig ist es für die Teilnehmenden, die betreffende Einstellung auf Instagram selbst zu finden. Die meisten können dies nur mit Hilfe des Interviewers und nur mit sehr spezifischen Anweisungen tun. Eine Person gibt sogar an, dass es Sinn macht, diese Funktion zu verstecken, da sie [Instagram] *„natürlich was davon [haben], wenn die Leute die Werbung gucken.“* (ITN13).

Die Einstellungsoption zu Werbeanzeigen wird vergleichsweise schneller gefunden – insgesamt ohne größere Probleme oder Anstrengungen. Das Auffinden der Einstellungsoption zu Cookies fällt den meisten Teilnehmenden ebenfalls leicht. Die Mehrheit der Teilnehmenden findet die Einstellung ohne große Schwierigkeiten und auch ohne Hilfe des Interviewers innerhalb kurzer Zeit. Einige brauchen zumindest leichte Hilfe, einige wenige zeigen große Schwierigkeiten bei der Suche nach der Einstellungsoption. Eine teilnehmende Person weist bei der Durchsicht der drei Auswahlmöglichkeiten für das Setzen von Cookies direkt darauf hin, dass die Informationstexte zu lang und komplex sind: *„Das ist ein echt langer Text. [...] Selbst wenn ich forciert werde, habe ich keinen Bock mir das alles durchzulesen. In meinem normalen Alltag würde ich mir das nicht angucken“* (ITN01). Darüber hinaus wird deutlich, dass den Teilnehmenden nicht immer klar ist, was die Einstellungen explizit bedeuten oder welche Auswirkungen sie haben – insbesondere, wenn die beiden Schieberegler für *Unsere Cookies in anderen Apps und auf anderen Websites* und *Cookies von anderen Unternehmen* aktiviert oder deaktiviert werden. Außerdem ist den Teilnehmenden unklar, was der Unterschied zwischen den Einstellungen in der Option *Werbeanzeigen* und *Daten von Partnern über Ihre Aktivitäten* ist. Ebenfalls herrscht Unwissen über die Tatsache, dass einige Einstellungen von Instagram direkt am Endgerät vorgenommen werden können, wie z. B. das Ein- und Ausschalten von Trackingfunktionen.

Es wurde bereits beschrieben, dass die Teilnehmenden beim Thema Datenschutz eine Art Hoffnungslosigkeit oder Ohnmacht empfinden. Eine zentrale Frage, die allen Teilnehmenden gestellt wurde und die sie beantworten mussten, war, ob sie bereit wären, die Datensammlung mit einem einzigen Klick massiv einzuschränken. Die Antworten zeigen, dass die Teilnehmenden durchaus zwischen mehr Privatsphäre und Datenschutz durch den Stopp der Sammlung persönlicher Daten und der Anzeige personalisierter Inhalte abwägen. Dabei geben einige Teilnehmende zu, dass ihnen die personalisierte Anzeige von Inhalten zusagt und sie diese nicht missen wollen. Eine Person führt zudem aus, es gut zu finden, dass der eigene Kanal durch den Algorithmus auch anderen Nutzern vorgeschlagen wird: *„[...] ich [mag] das eigentlich, wenn auch mal Fremde oder sonst was auf mein Profil stoßen [...]“* (ITN03).

Nur sieben der 20 Teilnehmenden geben an, dass sie von einer solchen Möglichkeit Gebrauch machen würden. Die Beweggründe hierfür sind recht unterschiedlich. Einige haben die Einstellung, dass sie, wenn sie bestimmte Inhalte sehen möchten, die entsprechenden Kanäle abonnieren. Bei Inhalten von Kanälen, denen sie nicht folgen, besteht nicht das Bedürfnis, diese sehen zu wollen:

*„Es geht mir hauptsächlich darum, was ich aktiv von den Leuten sehe, denen ich folge und nicht denen ich nicht folge.“ (ITN18)*

*„Ich kann ja die Seiten, die mich interessieren, abonnieren. Das wird mir dann angezeigt. Das drumherum empfinde ich dann eher als Locken. Von daher bin ich da überhaupt nicht böse, wenn mir das nicht mehr angezeigt wird.“ (ITN17)*

Für andere sei es einfach ein Mittel, um von Instagram wegzukommen, weil sie den Dienst viel zu intensiv nutzen, wie diese Person zum Beispiel angibt: *„[...] weil ich Instagram und den Inhalten auf Instagram persönlich nicht so eine große Bedeutung zuspreche. Es [wäre] [...] noch besser, wenn ich sie noch unansprechender fände, da ich mich selber manchmal erwische, dass ich dann doch mehr Zeit auf Instagram verbringe. Also ich fände es nicht schlecht, wenn ich es unansprechender fände.“ (ITN13).* Dies impliziert jedoch, dass der Wegfall der Personalisierung durch eine solche Maßnahme dazu beitragen würde, den Dienst nicht mehr oder zumindest weniger zu nutzen.

Die übrigen Teilnehmenden verneinen, dass sie eine solche Funktion nutzen würden, vor allem, wenn sie dafür auf Funktionalität, einschließlich Personalisierung, verzichten müssten:

*„Ich weiß es nicht, es kommt darauf an, wie sich das dann auswirken würde nachher. Ob [...] die eigene Benutzung nachher so eingeschränkt ist, dass man natürlich auch nicht mehr so das sieht, was man gerne möchte.“ (ITN04)*

*„Ich weiß nicht, weil eigentlich ist es ja für mich auch gut, wenn mir Sachen angezeigt werden, die ich gut finde. Ich will ja auch nicht, dass dann auf dieser Explorer Page irgendwelche Sachen über Werkzeuge [angezeigt werden], die mich überhaupt nicht interessieren.“ (ITN06)*

Die letztere Teilnehmende formuliert es nach der Durchsicht des Datenexports sogar wie folgt: *„[...] Im ersten Moment war ich ziemlich schockiert. Und dachte vielleicht sollte man sich doch mal lieber Gedanken über Cookies und all sowas machen. Aber im Endeffekt wiegte es doch mehr, als dass ich jetzt bereue, dass ich Instagram habe und diese ganzen Informationen über mich gespeichert sind.“ (ITN06)*

### 5.3 Verhaltensänderungen

Im zweiten Interview ging es vor allem darum, zu beobachten, ob sich nach dem ersten Interview und den beiden Interventionen das Verhalten in Bezug auf Instagram im Allgemeinen und den dortigen Datenschutz verändert hat, oder ob sich sogar die Datenschutzstrategie der Teilnehmenden auf der Plattform gewandelt hat. Es wird deutlich, dass die Teilnehmenden ganz unterschiedlich mit den Erkenntnissen aus dem ersten Interview umgegangen sind.

Bei insgesamt sechs Teilnehmenden lösen das erste Interview und die Interventionen keine Folgereaktion aus. Danach setzen sie sich weder mit den Themen Datenschutz und Algorithmen im Allgemeinen noch in Bezug auf Instagram auseinander. Es können auch keine Veränderungen in der Wahrnehmung von Instagram und den ausgespielten Inhalten festgestellt werden. Ebenfalls ändert diese Gruppe keine Einstellungen auf Instagram.

Die zweite Gruppe, bestehend aus neun Teilnehmenden, hat sich zumindest intensiver mit den im ersten Interview besprochenen Themen auseinandergesetzt. Einige von ihnen sind so weit gegangen, dass sie sich die Einstellungen angesehen und eigene Recherchen im Internet angestellt haben. Obwohl die meisten von ihnen nicht behaupten, irgendwelche Einstellungen geändert zu haben, sind sie sich der Inhalte, die ihnen auf Instagram angezeigt wurden, bewusster und versuchen zu verstehen, warum sie ihnen vorgeschlagen werden.

Die letzte Gruppe, bestehend aus fünf Teilnehmenden, beschäftigt sich in der Regel ausführlicher mit dem Thema und nimmt Änderungen an den Einstellungen vor oder probiert diese zumindest aus. Ein Teilnehmender entfernt zum Beispiel nach dem Interview alle Standardeinstellungen, die zu personalisierten Inhalten führen können, macht dies aber wieder rückgängig, nachdem ihm die angezeigten Inhalte nicht mehr gefallen. Andere Teilnehmende passen beispielsweise die Cookie-Einstellungen an, wechseln gelegentlich zum chronologischen Feed oder setzen zumindest das Konto auf privat. Eine dieser Personen führt aus: *„Ich habe mich tatsächlich noch mal mit den Cookie-Einstellungen beschäftigt. Ich habe mir das noch mal so ein bisschen durchgelesen, weil das war für mich tatsächlich so eine Überraschung, dass es das gibt – in so einer Social Media App.“* (ITN09). Dieselbe Person versucht auch, den Algorithmus zu steuern. Dafür sucht sie nach Themen und sieht sich Beiträge an, die nicht ganz dem eigenen Interesse entsprechen.

Die Teilnehmenden, die sich nicht weiter mit dem Thema befasst haben bzw. sich nicht mit den Einstellungen befasst oder gar etwas eingestellt haben, begründen dies schlicht mit mangelnder Zeit oder Notwendigkeit, dies zu tun. Ihre Datenschutzstrategie, oder das Fehlen einer solchen, hat sich für sie persönlich noch nicht negativ ausgewirkt. Die nähere Beschäftigung mit dem Thema auf dieser Ebene hat für sie also keine Priorität und sie fühlen sich dazu nicht motiviert.

Insgesamt wird deutlich, dass das erste Interview mit den beiden Interventionsmaßnahmen die Teilnehmenden zum Nachdenken anregt, auch wenn nur wenige von ihnen letztlich die Einstellungsmöglichkeiten wahrnehmen. Erwähnenswert ist auch, dass einige der Teilnehmenden beschrieben, ihr neu gewonnenes Wissen auch auf andere Online-Webseiten und Apps anzuwenden, indem sie zum Beispiel Cookie-Anfragen nicht immer gedankenlos akzeptieren.

Die Teilnehmenden sehen den Grund, warum sie sich selbst oder ihre Generation allgemein nicht intensiver mit dem Thema Datenschutz auseinandersetzen bzw. die Möglichkeiten, die in den Einstellungen angeboten werden, nicht nutzen, in einer Kombination aus Zeitmangel, Desinteresse und fehlendem Wissen. Eine Person erklärt dies zum Beispiel wie folgt: *„Also ich denke, erstens mal ist es einfach uferlos. Also wenn du anfängst dich mit Datenschutz zu beschäftigen, bei der Zahl von Social Media Accounts, die man heutzutage hat, dann kriegt man schon die Krise. Sobald man damit anfängt, ist es auch viel zu undurchsichtig, als dass man es irgendwie als Laie so schnell verstehen kann. Dann ist es auch ein bisschen unproduktiv [...]“* (ITN01). Ähnlich beschreibt es auch die folgende Person: *„Also so ein bisschen undurchsichtig ist das Thema. Es gibt sehr viel, was man dazu lesen muss, wenn man sich damit beschäftigt. Ich finde es teilweise auch schwierig formuliert, sodass man es [...] nicht direkt auf Anhieb versteht.“* (ITN09). Eine andere Person beschreibt in diesem Kontext das Gefühl, mit der Problematik allein gelassen zu werden: *„[...] man muss sich relativ eigenständig damit auseinandersetzen [...]. Man muss oft selber auf irgendwelche Felder klicken, damit man mehr [Informationen] angezeigt bekommt und dann ziemlich lange Texte durchlesen. Dann ist es natürlich wesentlich einfacher, nur ein Häkchen zu setzen und sich damit dann gar nicht auseinanderzusetzen.“* (ITN02). Ferner wird ausgeführt, dass die Existenz von Einstellungsmöglichkeiten meist nicht bekannt ist, der Zugang zu ihnen schwierig ist und mögliche Folgen von algorithmischen Entscheidungen und der damit verbundenen Datensammlung von Unternehmen für den Einzelnen nicht direkt spürbar und abschätzbar sind. Letzteres geht Hand in Hand mit den Antworten der Teilnehmenden auf die Frage, was passieren müsse,



damit sie sich mehr mit dem Thema und der Einstellung zum Datenschutz beschäftigen. Hier wird häufig angegeben, dass sie die negativen Auswirkungen spüren müssten. Die Bedenken hinsichtlich des Diebstahls und der Veröffentlichung privater Daten, wie Fotos oder Nachrichten, werden mehrfach geäußert. Schließlich gibt es einige Aussagen, die darauf hinweisen, dass man, wenn man einen Dienst nutzen möchte, keine andere Wahl hat, als die jeweiligen Datenschutzbestimmungen zu akzeptieren.

## 6 Schlussfolgerung

Algorithmen agieren in der Regel ohne das Bewusstsein vieler, die von ihren Entscheidungen betroffen sind. Dennoch ist ihr Einsatz weit verbreitet. Sie werden intern von Unternehmen und Institutionen genutzt, aber auch Einzelpersonen sind im Internet ständig mit algorithmischen Entscheidungen konfrontiert. Im Vergleich zu den anderen Anwendungsbereichen zeigen die vorgestellten Studien, dass die Verbraucher algorithmische Systeme im Internet zunehmend wahrnehmen. Doch obwohl auch die Teilnehmenden dieser Studie wissen, dass es automatisierte Prozesse gibt, die ihnen personalisierte Inhalte, Werbung und Empfehlungen liefern, zeigen die Ergebnisse der Interviewstudie, dass das Bewusstsein und das Wissen über die Funktionsweise von Algorithmen nicht oder nur oberflächlich vorhanden ist.

Die Teilnehmenden haben nur rudimentäres Wissen darüber, wie der für diese Studie ausgewählte Instagram-Algorithmus funktioniert und welche Daten dazu gesammelt und verwendet werden. Dies ermöglicht es ihnen nicht, fundiertere Entscheidungen darüber zu treffen, wie sie sich selbst schützen oder wie sie mit den algorithmischen Entscheidungen umgehen können und welche Auswirkungen die Entscheidungen auf sie haben. Je nachdem, um welche spezifischen Entscheidungen es sich handelt, kann dies dafür sorgen, dass diese blind angenommen werden, ohne sich möglicher Konsequenzen bewusst zu sein. Auch Diskriminierung kann dadurch nicht wahrgenommen werden oder wird akzeptiert.

Die neuen Anforderungen des Digital Services Act (DSA) und in Zukunft womöglich die des Gesetzes über künstliche Intelligenz zur Schaffung von mehr Transparenz für die Betroffenen können diesem Problem entgegenwirken. Die Ergebnisse dieser Arbeit und früherer Studien zeigen jedoch, dass Transparenz den Verbrauchern nicht unbedingt die aktive Kontrolle zurückgibt. Die durchgeführten Interviews und Interventionsmaßnahmen, die auch das Ziel von mehr Transparenz verfolgten, haben nur bei einem Teil der Teilnehmenden zu einem aktiven Umdenken und Handeln geführt, insbesondere im Hinblick auf ihre Einstellung zum Datenschutz. Die Grundstimmung, dass das Thema Datenschutz und die Art und Weise der Datenerhebung und -weiterverwendung für einen Laien zu komplex und aufwändig ist, um sie vollständig zu erfassen, hält sich auch nach der Intervention. Betont wird dabei ebenfalls, dass der Zugang zu den Einstellungen schwierig ist und mögliche Folgen von algorithmischen Entscheidungen und der damit verbundenen Datenerhebung durch Unternehmen für den Einzelnen nicht direkt spürbar und nicht abschätzbar sind.

Insgesamt werden daher neben mehr Transparenz auch Mechanismen benötigt, die Verbraucher anreizen und befähigen, sich mit Datenschutz auseinanderzusetzen und eine eigene Datenschutzstrategie zu entwickeln und im Alltag konsistent umzusetzen. Dies könnten nach Angaben der Teilnehmenden in dieser Studie folgende Ansätze sein:

- Intuitiveres Design der Oberfläche zu Datenschutzeinstellungen
- Pointierte, kurze Informationen
- Zentralisierte Einstellungsfunktionen bzw. einheitlicher Aufbau der Datenschutzeinstellungen (Standardisierung)
- Ausführliche Datenschutzabfrage bei der Installation von Apps
- Opt-in-Design

## Literaturverzeichnis

- Ali, Muhammad; Sapiezynski, Piotr; Bogen, Miranda; Korolova, Aleksandra; Mislove, Alan & Rieke, Aaron (2019). Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), S. 1-30.
- Angwin, Julia; Larson, Jeff; Mattu, Surya & Kirchner, Lauren (2016). Machine Bias. ProPublica. Abgerufen von: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Abgerufen am: 08.09.2022.
- Angwin, Julia; Scheiber, Noam & Tobin, Ariana (2017a). Machine Bias - Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads. ProPublica. Abgerufen von: <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>. Abgerufen am: 19.09.2022.
- Angwin, Julia; Tobin, Ariana & Varner, Madeleine (2017b). Machine Bias - Facebook (Still) Letting Housing Advertisers Exclude Users by Race. ProPublica. Abgerufen von: <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>. Abgerufen am: 21.09.2022.
- Araujo, Theo; Helberger, Natali; Kruike-meier, Sanne & de Vreese, Claes H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35 (3), S. 611-623.
- Arias-Cabarcos, Patricia; Khalili, Saina & Strufe, Thorsten (2022). 'Surprised, Shocked, Worried': User Reactions to Facebook Data Collection from Third Parties. *arXiv*. Abgerufen von: <https://arxiv.org/abs/2209.08048>. Abgerufen am: 20.12.2022.
- Bemmann, Florian; Windl, Maximiliane; Erbe, Jonas; Mayer, Sven & Husmann, Heinrich (2022). The Influence of Transparency and Control on the Willingness of Data Sharing in Adaptive Mobile Apps. *Proceedings of the ACM on Human-Computer Interaction*, 6 (MHCI), S. 1-26.
- Bigman, Y. E.; Wilson, D.; Arnestad, M. N.; Waytz, A. & Gray, K. (2022). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General* (Vorzeitige Online-Veröffentlichung).
- Blodgett, Su & O'Connor, Brendan (2017). Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English. *arXiv*. Abgerufen von: <https://arxiv.org/abs/1707.00061>. Abgerufen am: 19.09.2022.
- Bogert, Eric; Schecter, Aaron & Watson, Richard T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports*, 11 (1), S. 8028.
- Bonezzi, Andrea & Ostinelli, Massimiliano (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied*, 27 (2), S. 447-459.
- Bordalo, Pedro; Coffman, Katherine; Gennaioli, Nicola & Shleifer, Andrei (2016). Stereotypes. *The Quarterly Journal of Economics*, 131 (4), S. 1753-1794.
- Bryman, Alan (2003). *Quantity and Quality in Social Research*. London, GB: Routledge.
- Bucher, Taina (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20 (1), S. 30-44.
- Buolamwini, Joy & Gebru, Timnit (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, New York, NY, USA. *Proceedings of Machine Learning Research: PMLR*, 81, S. 77-91.
- Burrell, Jenna (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. In: *Big Data & Society*, 3 (1), S. 1-12.
- Castelluccia, Claude & Le Métayer, Daniel (2019). Understanding algorithmic decision-making: Opportunities and challenges. Panel for the Future of Science and Technology (STOA), European Parliamentary Research Service (EPRS), European Parliament. Brüssel, Belgien.
- Cofone, Ignacio N. (2019). Algorithmic Discrimination Is an Information Problem. *Hastings Law Journal*, 70 (6), S. 1389-1444.
- Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Roland L. & Stein, Clifford (2009). *Introduction to Algorithms*. 3. Cambridge, USA: The MIT Press.
- Cowgill, Bo; Dell'Acqua, Fabrizio; Deng, Samuel; Hsu, Daniel; Verma, Nakul & Chaintreau, Augustin (2020). Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics. *Proceedings of the 21st ACM Conference on Economics and Computation*, Virtuelles Event, Ungarn, S. 679-681.
- Cowgill, Bo & Tucker, Catherine E. (2019). *Algorithmic Fairness and Economics*. Columbia Business School Research Paper. Columbia University, New York, NY, USA.

- Datta, Amit; Tschantz, Michael Carl & Datta, Anupam (2015). Automated Experiments on Ad Privacy Settings. Proceedings on Privacy Enhancing Technologies, 2015 (1), S. 92-112.
- Diakopoulos, Nicholas (2014). Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center for Digital Journalism, Columbia University.
- Dietvorst, B. J.; Simmons, J. P. & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144 (1), S. 114-126.
- Dodge, Jonathan; Liao, Q. Vera; Zhang, Yunfeng; Bellamy, Rachel K. E. & Dugan, Casey (2019). Explaining models: an empirical study of how explanations impact fairness judgment. Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Rey, CA, USA, S. 275–285.
- Dogrue, Leyla; Facciorusso, Dominique & Stark, Birgit (2020). 'I'm still the master of the machine.' Internet users' awareness of algorithmic decision-making and their perception of its effect on their autonomy. Information, Communication & Society, 25 (9), S. 1311-1332.
- Egan, Paul & Roberts, Adrienne (2021). Judge: Companies can be sued over Michigan unemployment fraud fiasco. Detroit Free Press. Abgerufen von: <https://eu.freep.com/story/news/local/michigan/2021/03/26/judge-unemployment-midas-fraud-fast-enterprises-csg/7014975002/>. Abgerufen am: 30.11.2022.
- Eslami, Motahhare; Rickman, Aimee; Vaccaro, Kristen ; Aleyasen, Amirhossein; Vuong, Andy; Karrie, Karahalios.; Hamilton, Kevin & Sandvig, Christian (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about Invisible Algorithms in News Feeds. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republik Korea (Südkorea), S. 153–162
- Europäische Kommission (2022). Für ein besseres Internet: Das Gesetz über die digitalen Dienste (DSA) tritt heute in Kraft. Abgerufen von: [https://germany.representation.ec.europa.eu/news/fur-ein-besseres-internet-das-gesetz-uber-die-digitalen-dienste-dsa-tritt-heute-kraft-2022-11-16\\_de](https://germany.representation.ec.europa.eu/news/fur-ein-besseres-internet-das-gesetz-uber-die-digitalen-dienste-dsa-tritt-heute-kraft-2022-11-16_de). Abgerufen am: 13.12.2022
- Farke, Florian M.; Balash, David G.; Golla, Maximilian; Dürmuth, Markus & Aviv, Adam J. (2021). Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity. Proceedings of the 30th USENIX Security Symposium, Virtuelles Event, S. 483-500.
- Gran, Anne-Britt; Booth, Peter & Bucher, Taina (2021). To be or not to be algorithm aware: a question of a new digital divide? Information, Communication & Society, 24 (12), S. 1779-1796.
- Grzymek, Viktoria & Puntschuh, Michael (2019). What Europe Knows and Thinks About Algorithms - Results of a Representative Survey. Bertelsmann Stiftung. Gütersloh, Deutschland.
- Hannák, Anikó; Wagner, Claudia; Garcia, David; Mislove, Alan; Strohmaier, Markus & Wilson, Christo (2017). Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, OR, USA, S. 1914–1933.
- Hannák, Anikó; Wagner, Claudia; Garcia, David; Strohmaier, Markus & Wilson, Christo (2016). Bias in online freelance marketplaces: Evidence from taskrabbit. Proceedings DAT Workshop, Portland, OR, USA, S. 1914-1933.
- Herder, Eelco & van Maaren, Olaf (2020). Privacy Dashboards: The Impact of the Type of Personal Data and User Control on Trust and Perceived Risk. Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, S. 169-174.
- Kay, Matthew; Matuszek, Cynthia & Munson, Sean A. (2015). Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republik Korea (Südkorea) S. 3819–3828.
- Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil & Sunstein, Cass R (2019). Discrimination in the Age of Algorithms. In: Journal of Legal Analysis, 10, S. 113-174.
- Krempf, Stefan (2022). KI-Regeln: Über 40 Organisationen fordern Verbot von Predictive Policing. heise online. Abgerufen von: <https://www.heise.de/news/KI-Regeln-Ueber-40-Organisationen-fordern-Verbot-von-Predictive-Policing-6530587.html#:~:text=von%20Predictive%20Policing-,KI%2DRegeln%3A%20%C3%9Cber%2040%20Organisationen%20fordern%20Verbot%20von%20Predictive%20Policing,in%20der%20Strafverfolgung%20zu%20untersagen>. Abgerufen am: 06.10.2022.
- Lambrecht, Anja & Tucker, Catherine E. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. Management Science, 65 ( 7), S. 2966-2981.
- Langer, Markus; König, Cornelius J.; Back, Caroline & Hemsing, Victoria (2022). Trust in Artificial Intelligence: Comparing Trust Processes Between Human and Automated Trustees in Light of Unfair Bias. Journal of Business and Psychology, 2022.

- Leslie, David; Mazumder, Anjali; Peppin, Aidan; Wolters, Maria K. & Hagerty, Alexa (2021). Does “AI” stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*, S. 372-376.
- Logg, Jennifer M.; Minson, Julia A. & Moore, Don A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, S. 90-103.
- Lum, Kristian & Isaac, William (2016). To predict and serve? *Significance*, 13 (5), S. 14-19.
- Mehrabi, Ninareh; Morstatter, Fred; Saxena, Nripsuta; Lerman, Kristina & Galstyan, Aram (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Survey*, 54 (6), S. 1-35.
- Mohajan, Haradhan Kumar (2018). Qualitative Research Methodology in Social Sciences and Related Subjects. *Journal of Economic Development, Environment and People*, 7 (1), S. 23-48.
- Obermeyer, Ziad; Powers, Brian; Vogeli, Christine & Mullainathan, Sendhil (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366 (6464), S. 447-453.
- Orwat, Carsten (2019). Risks of Discrimination through the Use of Algorithms. Federal Anti-Discrimination Agency. Berlin, Deutschland.
- Patton, Michael Quinn (2002). *Qualitative Research & Evaluation Methods*. Bd. 3. Thousand Oaks, CA, USA: Sage Publications.
- Powers, Elia (2017). My News Feed is Filtered? *Digital Journalism*, 5 (10), S. 1315-1335.
- Prates, Marcelo O. R.; Avelar, Pedro H. & Lamb, Luís C. (2020). Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 32 (10), S. 6363-6381.
- Rader, Emilee (2014). Awareness of behavioral tracking and information privacy concern in Facebook and Google. *Proceedings of the 10th USENIX Conference on Usable Privacy and Security*, Menlo Park, CA, USA, S. 51-67.
- Rader, Emilee; Cotter, Kelley & Cho, Janghee (2018). Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Kanada, S. 1-13.
- Rader, Emilee & Gray, Rebecca (2015). Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republik Korea (Südkorea), S. 173–182.
- Rao, Ashwini; Schaub, Florian & Sadeh, Norman (2014). What do they know about me? Contents and Concerns of Online Behavioral Profiles. *PASSAT '14: 6th ASE International Conference on Privacy, Security, Risk and Trust*, Cambridge, MA, USA.
- Renier, Laetitia A.; Schmid Mast, Marianne & Bekbergenova, Anely (2021). To err is human, not algorithmic – Robust reactions to erring algorithms. *Computers in Human Behavior*, 124, S. 106879.
- Schwartzstein, Joshua (2014). Selective Attention and Learning. *Journal of the European Economic Association*, 12 (6), S. 1423-1452.
- Siles, Ignacio; Espinoza-Rojas, Johan; Naranjo, Adrián & Tristán, María Fernanda (2019). The Mutual Domestication of Users and Algorithmic Recommendations on Netflix. *Communication, Culture and Critique*, 12 (4), S. 499-518.
- Snow, Jacob (2018). Amazon’s Face Recognition Falsely Matched 28 Members of Congress With Mugshots. ACLU. Abgerufen von: <https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28>. Abgerufen am: 15.09.2022.
- Statista (2022). Umfrage zur Nutzung von Instagram nach Altersgruppen in Deutschland bis 2022. Abgerufen von: <https://de.statista.com/statistik/daten/studie/691584/umfrage/anteil-der-nutzer-von-instagram-nach-alter-in-deutschland/>. Abgerufen am: 15.07.2022.
- Sundar, S. Shyam (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In: Metzger, M.; Flanagin, A.: *Digital Media, Youth, and Credibility*. Cambridge, MA, USA: MIT Press.
- Swart, Joëlle (2021). Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media. *Social Media & Society*, 7 (2), S. 20563051211008828.
- Valentino-DeVries, Jennifer; Singer-Vine, Jeremy & Soltani, Ashkan (2012). Websites Vary Prices, Deals Based on Users' Information. *The Wall Street Journal*. Abgerufen von: <https://www.wsj.com/articles/SB10001424127887323777204578189391813881534>. Abgerufen am: 25.11.2022.
- Yeomans, Michael; Shah, Anuj; Mullainathan, Sendhil & Kleinberg, Jon (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32 (4), S. 403-414.