

Computer- und Informationsethik

Helena Häußler*

Data Ethics Frameworks

Werte und Machtstrukturen im Spannungsfeld von Technologie und moralischem Handeln

<https://doi.org/10.1515/iwp-2021-2178>

Zusammenfassung: Zuletzt veröffentlichten viele Organisationen ethische Richtlinien, um ihre Haltung gegen Diskriminierung durch Algorithmen zu betonen. Vier dieser Frameworks werden mithilfe der kritischen Diskursanalyse untersucht. Ziel ist es, die darin vermittelten Werte und Wertkonflikte zu identifizieren. Die Ergebnisse weisen darauf hin, dass etablierte Werte aus der Computer- und Informationsethik aufgegriffen und bestehende Machtstrukturen zwischen Akteuren verstärkt werden.

Deskriptoren: Ethik, Informationsgesellschaft, Informationstechnologie, Richtlinie, Linguistik, Data Ethics Framework, Werte, Computer- und Informationsethik, Diskursanalyse

Data Ethics Frameworks

Values and power structures in tension between technology and moral conduct

Abstract: Recently, many organizations published ethical guidelines to emphasize their stance against algorithmic bias. Four of these frameworks are examined using Critical Discourse Analysis. The aim is to identify the values and value conflicts conveyed in them. The results indicate that established values from computer and information ethics are taken up and existing power structures between actors are reinforced.

Descriptors: Data Ethics Framework, Values, Computer and Information Ethics, Critical Discourse Analysis, Information Society, Information Technology, Policy, Linguistics

Cadres d'éthique numérique

Valeurs et structures de pouvoir dans le domaine de la tension entre technologie et action morale

Résumé: Plus récemment, de nombreuses organisations ont publié des directives éthiques pour souligner leur position contre la discrimination par les algorithmes. Quatre cadres sont examinés à l'aide d'une analyse critique du discours. L'objectif est d'identifier les valeurs et leurs conflits qu'elles véhiculent. Les résultats indiquent que les valeurs établies de l'éthique de l'information sont reprises et que les structures de pouvoir existantes sont renforcées.

Descripteurs: Cadres d'éthique numérique, Valeurs, L'éthique d'information, Analyse critique du discours, Éthique, société de l'information, technologie de l'information, politique, linguistique

Einführung und Problemstellung

Im Sommer 2020 gingen britische Jugendliche auf die Straße, um gegen den Algorithmus zu protestieren, der zur Berechnung der Abschlussnoten eingesetzt worden war. Anstelle der Einschätzungen der Lehrkräfte sollte ein automatisiertes Entscheidungssystem die Noten errechnen, bevorzugte dabei aber Privatschulen, die in der Vergangenheit bessere Abschlüsse vorgewiesen hatten (Holland 2020). Dies ist nur ein aktuelles Beispiel in einer Reihe von Fällen von *algorithmic bias* – Algorithmen, die Individuen oder Gruppen systematisch und unfair benachteiligen (Friedman und Nissenbaum 1996).

Unternehmen und öffentliche Einrichtungen reagieren auf diese Fälle üblicherweise mit Korrekturen der Systeme – und zunehmend mit der Betonung von wertorientiertem Handeln. Dies lässt sich an den Richtlinien und Kodizes ablesen, die vermehrt nach 2016 veröffentlicht wurden und das Verhalten der Angestellten an ethischen Prinzipien ausrichten sollen (Whittlestone et al. 2019b). Unverbindliche Statements stellen sich jedoch als wenig hilfreich für die tatsächliche Anwendung heraus, sodass zuletzt auch praxisnahe und niedrigschwellige Instru-

*Kontaktperson: Helena Häußler, Hochschule für Angewandte Wissenschaften (HAW) Hamburg, Fakultät Design, Medien & Information, Department Information, Finkenau 35, 22081 Hamburg, E-Mail: helena.haeussler@haw-hamburg.de

mente entwickelt wurden, die *Data Ethics Frameworks*. Diese sollen in der Design- und Entwicklungsphase von Technologien und Anwendungen zur Reflektion der ethischen Implikationen eingesetzt werden. Daher richten sie sich nicht nur an technisches Personal, sondern an alle, die in dem jeweiligen Projekt eingesetzt sind (z. B. DataEthics.eu 2021; Schäfer und Franzke 2020; Tarrant und Madison 2021).

Die Frameworks bilden so einen Indikator dafür, wie die Auseinandersetzung um die (ethische) Gestaltung des technischen Fortschritts dargestellt wird. In der Masterarbeit, auf der dieser Beitrag beruht, wurden vier ausgewählte Data Ethics Frameworks auf die von ihnen vermittelten Werte und Wertkonflikte untersucht. Ziel ist die thematischen Prioritäten und Machtstrukturen darzustellen und vor dem aktuellen Hintergrund zu hinterfragen. Dafür eignet sich die kritische Diskursanalyse nach Norman Fairclough, mit der Sprache in konkreten Texten ideologiekritisch darauf geprüft wird, wie sie aktuelle Hegemonie abbildet oder zu sozialem Wandel beiträgt. Dabei werden die Texte der Frameworks im Kontext ihrer Entstehung und ihres sozialen Umfelds betrachtet. Diese holistische Perspektive unterscheidet die vorliegende Arbeit von bisherigen Analysen ethischer Guidelines (Jobin et al. 2019; Hagendorff 2020; Schiff et al. 2021).

Im Folgenden wird der soziale Kontext anhand historischer und aktueller Themen in der Disziplin der Computer- und Informationsethik (CIE) erläutert. Des Weiteren wird auf die Methodik der kritischen Diskursanalyse eingegangen und die Auswertung beschrieben. Die identifizierten Werte, Wertkonflikte und Machtstrukturen werden mithilfe von Erkenntnissen aus der Analyse illustriert. Abschließend werden einige Aspekte vertieft diskutiert, die aus sozialkonstruktivistischer Sicht relevant sind, bevor ein abschließendes Fazit gezogen wird.

Überblick Computer und Informationsethik

Straftaten, die zunehmend mithilfe von Computern und von Computerfachleuten durchgeführt wurden, gaben den Anstoß für die Anfänge der Computer- und Informationsethik (CIE) (Bynum 2010). Moor (1985) führt dies auf die logische Formbarkeit von Computern zurück, mit denen Aktivitäten durchgeführt werden können, die vorher nicht möglich waren. Menschen können sich daher bei der Benutzung nicht an etablierten Handlungsweisen und Normen orientieren, sodass ein *policy vacuum* entsteht (Moor 1985, S. 266). Den Auftrag der Computerethik sieht

Moor in der konzeptionellen Analyse möglicher ethischer Probleme und in der Ableitung von moralisch richtigen Verhaltensweisen, die auch Anhaltspunkte für die Gesetzgebung bieten.

Seither hat die Disziplin sich analog zum technischen Fortschritt weiterentwickelt. Die Ursachen für auftretende ethische Fragestellungen werden heute weniger in Computern verortet, sondern im Umgang mit Informationen und auf einem anderen Abstraktionsniveau auf der Ebene von Daten (Floridi und Taddeo 2016), wenngleich sich die Begrifflichkeiten zuletzt uneinheitlich ausdifferenziert haben (Whittlestone et al. 2019b). An einer aktuellen Definition von Data Ethics zeigt sich jedoch, dass sich seit Moor (1985) wenig an der Auffassung der Disziplin geändert hat: “Data Ethics can be defined as the branch of ethics that studies and evaluates moral problems related to data [...], algorithms [...], and corresponding practices [...], in order to formulate and support morally good solutions (e. g. right conducts or right values)” (Floridi und Taddeo 2016, S. 3).¹

Diese Einteilung spiegelt sich auch in der aktuellen Forschung wider. *Algorithmic bias* ist ein breit behandelter Aspekt, aber auch die Verletzung von ethischen Prinzipien wie Autonomie und Privatsphäre in konkreten Fällen gehören zu den untersuchten Problemen (Ananny 2016; Mittelstadt et al. 2016; Saltz und Dewar 2019). Frühe Ansätze zur Lösung von ethischen Problemen waren Bildung und ethische Richtlinien – beides wird weiterhin verfolgt (Shapiro et al. 2020). Zunehmend werden technische Optimierungsmaßnahmen zur Vermeidung von ethischen Problemen diskutiert, etwa die statistische Angleichung von Trainingsdaten für eine faire Repräsentation (Dunkelau und Leuschel 2019).

Die Bewahrung und Förderung von menschlichen Werten ist ein Hauptanliegen der CIE. Werte werden in dem Zusammenhang verstanden als Ideale menschlichen Handelns untereinander sowie darüber, wie die Gesellschaft organisiert sein sollte, um moralisch richtiges Handeln zu unterstützen (Brey 2010). Während anfangs versucht wurde, Kernwerte zu definieren, sind heutige Ansätze vielfältiger (La Fors et al. 2019), wobei weiterhin kein Konsens darüber besteht, ob universale Prinzipien die menschliche Komplexität angemessen abbilden können (van den Hoven 2010). Ansätze wie *Value Sensitive Design* orientieren sich an Werten in dem Bewusstsein, dass diese

¹ Da sich die Bezeichnung bisher nicht durchgesetzt hat, wird hier Computer- und Informationsethik verwendet, um von der akademischen Disziplin zu sprechen.

bei der Entwicklung von Systemen und Anwendungen einfließen (Friedman und Hendry 2019). Diese Auffassung versteht Technologie nicht deterministisch, sondern erkennt den sozialen Einfluss an und ist demzufolge offen für menschliches Handlungs- und Veränderungspotential. Darin zeigen sich die Parallelen zum Sozialkonstruktivismus und zum Ansatz der kritischen Diskursanalyse, die nicht nur Methodik, sondern auch ein breites theoretisches Fundament bietet.

Methodik und Analyse

Der sozialkonstruktivistische Ansatz geht davon aus, dass das Wissen über die Welt und die Menschen nicht ausschließlich in einer externen, messbaren Realität gründet, sondern durch Kulturen, Normen und soziale Zusammenhänge geformt wird (Burr 2015). Viele theoretische Ansätze konzentrieren sich auf Sprache, da sich daran Interaktionen sowie die gegenseitige Konstruktion zwischen Text und sozialem Umfeld besonders deutlich zeigen. Die *kritische Diskursanalyse* nach Fairclough geht von der Annahme aus, dass Sprache ideologisch ist und sich darin die vorherrschende soziale Gruppe manifestiert (Fairclough 1995). Ein Diskurs bzw. eine Diskursordnung zeigen sich daran, wie Worte und Bedeutungen in einen stabilen Sinnzusammenhang gebracht und in der gleichen Form im gleichen Kontext verwendet werden (ebd.).

Faircloughs Rahmenmodell zeigt die Einbettung von Text in diskursive Praktiken der Erstellung, Verbreitung und Rezeption der Texte und im weitesten Kontext in die soziokulturellen Praktiken eines bestimmten sozialen Felds, die sich in Überzeugungen, Haltungen und Strukturen ausdrücken. Während sich Text und soziokulturelle Praktiken gegenseitig formen, ist die diskursive Praxis die ausschlaggebende Dimension, über die sozialer Wandel oder Stärkung des Bestehenden ermöglicht werden (Fairclough 2001). Das Handlungspotential der beteiligten Personen liegt darin, mit kreativer Verwendung von Sprache Widerstand zu leisten bzw. mit Rückgriff auf etablierte Prozesse und Formulierungen am Status quo festzuhalten. Der Diskursdimension nähert sich Fairclough über die linguistische Analyse der vorliegenden Texte, während die diskursiven Praktiken anhand einer Prozessanalyse interpretiert werden und die soziokulturellen Praktiken mithilfe von sozialer Theorie Erklärungen liefern (s. Abb. 1).

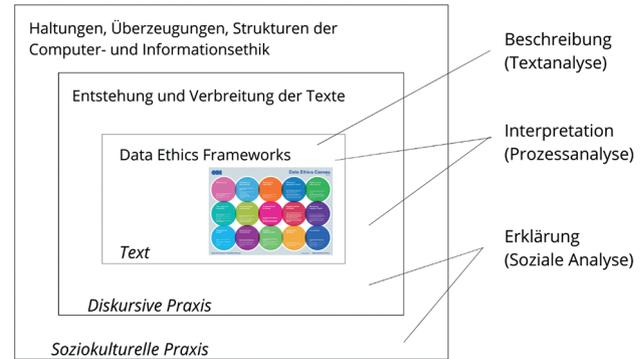


Abbildung 1: Faircloughs Modell von Diskurs (kursiv) und Diskursanalyse (rechts), angewandt auf die vorliegende Analyse. (Eigene Darstellung auf Grundlage von Fairclough 1995, S. 98)

Vier Data Ethics Frameworks² von öffentlichen oder gemeinnützigen Organisationen wurden für die Analyse ausgewählt:

- Data Ethics Decision Aid (DEDA) der Utrecht Data School
- Digital Decision Tool (DD Tool) des Center for Democracy and Technology
- Data Ethics Canvas (DEC) des Open Data Institute
- Data Ethics Workbook (DEW) des Departments für Digitales, Kultur, Medien & Sport des britischen Cabinet Office

Alle sind verzeichnet im „AI Ethics Guidelines Global Inventory“ von AlgorithmWatch³ und wurden zwischen 2016 und 2017 veröffentlicht. Die Auswahl dieser Frameworks begründet sich durch ihren frageförmigen Aufbau, sowie das praxisorientierte, innovative Design in einem vergleichbaren, knappen Umfang.

Für die linguistische Analyse wurden Transkripte der Frameworks erstellt und auf textliche Muster im Vokabular, Modalverben, Passivstrukturen und dargestellte Akteure untersucht (Fairclough 1989). Mithilfe der integrierten Wertliste von La Fors et al. (2019) wurden die Werte codiert sowie Wertkonflikte und präsenste Diskurse gekennzeichnet. Die diskursiven Praktiken um die Erstellung und Verbreitung der Frameworks wurden anhand der Informationen aus Begleitmaterialien und von den Webseiten bzw. Publikationen der Organisationen ausgewertet. Der Forschungsüberblick dient als Ausgangs-

² Die Frameworks sowie die Transkriptionen der Texte sind in ihrer verwendeten Version zugänglich unter: <https://osf.io/z7e9m/> [27.7.2022].

³ Siehe: <https://inventory.algorithmwatch.org/> [27.7.2021].

punkt für die Einordnung in die soziokulturellen Praktiken der Computer- und Informationsethik. Da es keinen Konsens in der Frage nach der Validität von Diskursanalyse gibt, wird Jørgensen und Phillips (2002) gefolgt, die für eine solide Analyse plädieren, basierend auf mehreren untersuchten Merkmalen sowie einem transparenten Umgang mit Widersprüchen in der Interpretation.⁴

Ergebnisse

In den vier Frameworks wurden insgesamt 646 Einheiten codiert, verteilt auf identifizierte Werte und Wertkonflikte, präsenste Diskurse und Akteure. Die integrierte Liste an Werten hat sich zum überwiegenden Teil bestätigt. Folgende Werte konnten festgestellt werden: *Wohl der Menschheit*, *Schadensvermeidung*, *Gerechtigkeit*, *Transparenz*, *Verantwortlichkeit*, *Autonomie* und *Privatsphäre*. Wohl der Umwelt war der einzige Wert, der in keinem Framework erwähnt wurde, wohingegen der Wert Transparenz aufgrund seiner Präsenz als separater Wert definiert wurde. Unter *Wohl der Menschheit* wird ein allgemeiner positiver Einfluss auf die Gesellschaft verstanden, hingegen sind konkret Missbrauch, Diskriminierung und Falschinterpretation als Aspekte der *Schadensvermeidung* genannt. Zusätzlich können das Vorsorgeprinzip und das Arbeiten innerhalb der eigenen Kompetenzen dazu dienen, negative Auswirkungen zu vermeiden. *Gerechtigkeit* zielt vor allem auf Diskriminierungsfreiheit ab, etwa durch gleiche Repräsentation in (Trainings-)Daten, sowie am Rande auf universelle Nutzbarkeit und Inklusion. Mit *Transparenz* wird Bezug genommen auf Erklärbarkeit und Offenheit mit Zielen und Methoden des Projekts. *Verantwortlichkeit* wird als persönliche Haftung dargestellt, sodass die Nachvollziehbarkeit einen großen Anteil einnimmt. Mit *Autonomie* meinen die Frameworks Handlungs- und Beteiligungsoptionen, die oft dargestellt werden als eingegrenzt durch gesetzliche Vorgaben, technische und organisationale Umstände. Bei der Einwilligung in die Einbindung in das Projekt finden sich Schnittstellen mit *Privatsphäre*, wobei darüber hinaus der Schutz vor Identifizierung oder der Umgang mit sensiblen Daten darunterfallen.

Im Vorkommen der Werte zeigen sich deutliche Unterschiede zwischen den Frameworks. DEDA geht detailliert auf Privatsphäre (n=8) ein, während bei DD Tool Gerech-

tigkeit (n=16) und Autonomie (n=10) im Vordergrund stehen und Wohl der Menschheit nicht vorkommt. DEC setzt einen Schwerpunkt auf Transparenz (n=11), DEW hingegen auf Verantwortlichkeit (n=11). Sprachliche Muster zeigen sich etwa dabei, dass ein positiver Mehrwert des Projekts im Präsens formuliert und negative Auswirkungen mit Modalverben ausgedrückt werden. Laut Fairclough (1989) zeigt sich daran die Wahrnehmung von Realität, also die Überzeugung, dass Technologie grundsätzlich einen positiven Beitrag leistet. Passivstrukturen kommen häufig vor, wenn die Fragen negative Auswirkungen betreffen, was eine Tendenz zum Technologiedeterminismus aufweist und die Ursache für eigentliche Probleme verschleiert.

Auffallend ist, dass nur DEDA einen allgemeinen Wertediskurs thematisiert, in dessen Rahmen die persönlichen Wertvorstellungen der Nutzenden des Frameworks eine Rolle spielen. Diese Anpassung erfolgte durch die Erfahrungen der Utrecht Data School aus Workshops mit Anwenderinnen und Anwendern des Frameworks (Siffels et al. Forthcoming 2021). Weitaus häufiger werden wirtschaftliche, technische und juristische Diskurse zur Darstellung der Werte bedient, die sich durch entsprechendes Fachvokabular äußern. Des Weiteren konnten auch ein Antidiskriminierungsdiskurs und ein Demokratisierungsdiskurs wahrgenommen werden, die vor allem darauf zielen, Diskriminierung zu beseitigen und Partizipation und Aufklärung zu steigern.

Zwischen den Frameworks zeigen sich deutliche Unterschiede in der Verwendung der Diskurse. Das DD Tool weist etwa vor allem einen technischen und Antidiskriminierungsdiskurs auf, während DEC am häufigsten einen wirtschaftlichen Diskurs anführt. Dies lässt sich auf den Entstehungskontext zurückführen: DD Tool entstand aus einem Projekt zu *algorithmic bias* und beschäftigte sich mit der technischen Perspektive (Lange 2016). Für DEC bildete ein Business Canvas Modell die Vorlage und zumindest in Begleitmaterialien war eine Firma aus dem Netzwerk der Organisation beteiligt (Osterwalder et al. 2014; Broad et al. 2017).

Wertkonflikte finden sich in den Frameworks in einem begrenzten Ausmaß. Auch an dieser Stelle hebt sich DEDA von den anderen Dokumenten ab, da darin Konflikte zwischen persönlichen und organisationalen Werten erwähnt werden. Weitere Konflikte zeigen sich vielmehr implizit, indem in einer Frage Interessen mehrerer Gruppen abgewogen oder Projektziele den Prioritäten von anderen Gruppen gegenübergestellt werden. Eine gezielte Thematisierung von Konflikten, von Werten oder den Wertvorstellungen verschiedener Personen findet nicht statt. Dafür fehlt die Grundlage, da Werte durchweg einzeln adres-

⁴ Die Analyse kann nachvollzogen werden anhand von Codebuch und codierten Segmenten, zugänglich unter: <https://osf.io/z7e9m/> [27.7.2021].

sirt werden und diese nicht in Beziehung oder Interaktion miteinander gesetzt werden.

Machtstrukturen zeigen sich in der Darstellung der Akteure in den Frameworks. Wenig überraschend überwiegen die Referenzen zu den im Design- und Entwicklungsprojekt beteiligten Personen, da es sich um die angesprochene Zielgruppe handelt, welche die Frameworks anwenden sollen. Dementsprechend wird auf die Werte, Stereotypen, Annahmen und Gefühle des Projektteams⁵ eingegangen, die hinterfragt und reflektiert werden sollen. Bei den weiteren Akteuren handelt es sich um eine heterogene Stakeholdergruppe, auf die Bezug genommen wird als allgemein Betroffene des Datenprojekts sowie Involvierte, Diskriminierte und Datensubjekte. Auf deren Werte, Weltanschauungen oder Gefühle wird, abgesehen von Diskriminierungserfahrungen, nicht eingegangen.

Während das Projektteam einen aktiv beschriebenen Handlungsspielraum nutzt, werden die Aktivitäten der Stakeholder vom Projektteam bestimmt, etwa der Grad der Beteiligung, und erschöpfen sich häufig in Reaktionen auf dessen Aktionen. Vorbehalte gegenüber den Stakeholdern werden sichtbar an der Sorge vor kritischen Reaktionen und Nutzung bzw. Interpretation abweichend von der intendierten. Parallelen zeigen sich zu den Prozessen um die Erstellung der Frameworks: die jeweiligen Organisationen übernehmen die federführende Rolle. Stichpunktartig bzw. bei der Überarbeitung der Frameworks wurden Personen aus der Zielgruppe oder die Bevölkerung mit einbezogen (Ginnis et al. 2016; Utrecht Data School 2020; Vryzakakis und Thereaux 2020).

Diskussion

Die in den Data Ethics Frameworks identifizierten Werte weisen große Übereinstimmungen mit vorherigen Auswertungen von ethischen Richtlinien auf (Jobin et al. 2019; Hagendorff 2020). Auch die wenig referenzierten Wertkonflikte zeigen die enge Verbindung zur Computer- und Informationsethik, da in der Hinsicht weiterhin eine große Forschungslücke besteht (Manders-Huits und Zimmer 2009; Whittlestone et al. 2019a). Selbst wenn die Form und der Aufbau der Frameworks sich von anderen Guidelines abheben, so sind doch die Inhalte wenig innovativ. Dies zeigt sich auf der Textebene an den etablierten Begrifflichkeiten wie Schadensvermeidung oder Privatsphä-

re. Im Gegensatz dazu ist bei aktuellen kritischen Ansätzen wie *Data Justice* (D'Ignazio und Klein 2020) oder *Good Data* (Trenham und Steer 2019) die disruptive Haltung bereits an der Formulierung erkennbar, die zudem an der Zusammensetzung und Definition der angeführten Werte deutlich wird. Darüber hinaus zeigt sich an dem Verständnis von Problemen und Lösungen, wie die Frameworks Ausdruck der akademischen Disziplin sind. Ethische Dilemmas werden dabei häufig in Bezug auf die Handlung oder die Auswirkungen betrachtet. Lösungen tendieren dazu, das Narrativ der „bösen Akteure“ zu bedienen oder versuchen, mithilfe von Optimierung der verwendeten Daten und Algorithmen die durch sie auftretenden Probleme zu verhindern. Diese Sichtweise ist dafür kritisiert worden, die zugrundeliegenden sozialen Umstände auszublenden und die historischen Bedingungen für die präsenten Verhältnisse nicht zu berücksichtigen (Hoffman 2019; Dencik et al. 2019).

Die Praktiken bei der Erstellung der Texte haben zu einer konventionellen Zusammenstellung von Werten beigetragen. Auch wenn die Organisationen interdisziplinär zusammengesetzt sind, ist nur eine kleine Meinungsvielfalt repräsentiert. Eine breite Beteiligung von Stakeholdern – denjenigen, die Anwendungen benutzen, deren Daten verwendet werden, die von den Auswirkungen der automatisierten Systeme betroffen oder benachteiligt sind – bei der Erstellung der Frameworks war nicht vorgesehen. In den Texten zeigt sich das daran, dass eine anlasslose Beteiligung kaum thematisiert wurde und Partizipation hauptsächlich vom Projektteam festgelegt wird. Dadurch werden große Gruppen von der Auseinandersetzung um die Gestaltung von technischem Fortschritt ausgeschlossen. An den Machtverhältnissen, die wenige Menschen in der Entwicklung von Anwendungen und Systemen begünstigt, ändert sich nichts (boyd und Crawford 2012). Ein Forschungsdesiderat in der Disziplin zeigt sich auch bei effektiven Methoden zur Beteiligung von verschiedenen Stakeholdergruppen (Friedman und Hendry 2019).

Die Analyse der Frameworks hat offenbart, dass kaum auf ethische Diskurse zurückgegriffen wird, um zur Reflexion von ethischen Auswirkungen anzuregen. Das Überwiegen der Präsenz anderer Diskurse, z. B. ein wirtschaftlicher Diskurs, die zur Konzeptionalisierung von Werten herangezogen wurden, weist auf eine Veränderung hin. Fairclough bezeichnet es als *hegemonic struggle*, wenn sich die Bedeutung von Worten durch einen neuen Kontext ändert, in dem sie interpretiert werden (Fairclough 2010). Die Kommerzialisierung des Gegenstands Data Ethics legt nahe, dass nicht die vorher unterlegene Gruppe durch den Machtkampf in eine andere Position kommt.

⁵ Darunter wird die Zielgruppe der Designer/innen, Entwickler/innen, Projektmanager/innen, Analyst/innen und Entscheidungsträger/innen zusammengefasst.

In der Computer- und Informationsethik bestehen vielfache Verschränkungen mit der Tech-Industrie. Forschungsergebnisse beschäftigen sich häufig mit den ethischen Problemen, die bei den Anwendungen und automatisierten Systemen der Firmen auftreten, sind bei der Erarbeitung von moralisch richtigen Verhaltensweisen jedoch darauf angewiesen, dass diese auch von der Industrie aufgegriffen werden (Mittelstadt 2019). Dies impliziert, dass die Argumentation für ethische Technologien nicht nur auf Prinzipien ansetzen kann, sondern auch mit einem Geschäftsvorteil den Aufwand wirkungsvoll begründen muss (Metcalf et al. 2019).

Der Einfluss der Tech-Industrie auf die Disziplin zeigt sich auch finanziell, etwa am Sponsoring für akademische Konferenzen, Finanzierung von Instituten – wie Facebooks Engagement an der TU München – und an der Unterhaltung von eigenen Teams mit anerkannten Wissenschaftlerinnen und Wissenschaftlern. Somit beteiligen sich große Firmen auch inhaltlich am akademischen Diskurs. Googles umstrittene Entlassung der AI Ethics Forscherin Timnit Gebru Ende 2020 demonstriert, dass kein konstruktiver Umgang mit kritischen Stimmen möglich ist. Gebru beschäftigte sich mit den Risiken von aktuellen Sprachmodellen und setzte sich offensiv mit ihrer Perspektive als schwarzer Frau in der KI-Forschung auseinander (Simonite 2021). Firmen, die sich für Data Ethics engagieren, ohne wirkungsvoll etwas an ihren Geschäftspraktiken zu ändern, stehen unter dem Verdacht des *Ethics Bluewashing* (Floridi 2019), wobei auch Forschungseinrichtungen sich dem Vorwurf der Auftragsforschung aussetzen (Ochigame 2019).

Angesichts der wenigen beteiligten Akteure, sich verfestigenden Haltungen und den Abhängigkeiten von der Industrie stellt sich die Frage, wie sich die Disziplin der Computer- und Informationsethik weiterentwickeln kann. Mit dem kontinuierlichen technischen Fortschritt sowie dem zunehmenden Einfluss von automatisierten Entscheidungssystemen in vielen gesellschaftlichen Bereichen bleiben Fragen nach moralisch richtigen Handlungsweisen aktuell und drängend. Die Disziplin ist sowohl abhängig von den Einblicken in die sich ebenfalls verändernden Wertvorstellungen der Menschen, die direkt oder indirekt im Kontakt mit den Systemen sind, als auch von der Akzeptanz und Anwendungsbereitschaft der Tech-Industrie.

Zusammenfassung und Ausblick

In dem vorliegenden Beitrag wurden mithilfe der kritischen Diskursanalyse vier Data Ethics Frameworks ausgewertet und auf die zugrundeliegenden Werte und Macht-

strukturen untersucht. Die identifizierten Werte sind in ihren Begrifflichkeiten und in ihrem Verständnis Ausdruck von etablierten, nicht ganz unkritischen Annahmen in der Computer- und Informationsethik. Dieser Eindruck wird unterstützt von der marginalen Abdeckung von Wertkonflikten, die sich auch in bestehenden Forschungslücken widerspiegelt. Veränderungen zeigen sich in Diskursen, in deren Kontext sich die Bedeutung der Werte verändert und weg von ethischen Grundsätzen hin zu technischen, wirtschaftlichen und juristischen Diskursen weist. Aus der Perspektive Faircloughs handelt es sich dabei jedoch nicht um eine Veränderung, die den sozialen Wandel der nicht herrschenden, unterdrückten Gruppe unterstützt. Vielmehr setzt sich so die Perspektive derer durch, die ohnehin großen Einfluss ausüben, weil sie den technischen Fortschritt maßgeblich vorantreiben und bestimmen. Die vorliegende Analyse soll dazu dienen, sich mit der Weiterentwicklung der Disziplin kritisch auseinanderzusetzen, um zukünftigen ethischen Problemen gerecht zu werden, ohne etablierte Strukturen unhinterfragt zu verfestigen. Erste Schritte können niedrigschwellig auf der diskursiven Ebene umgesetzt werden, indem verschiedene Stakeholder und ihre Sprache bei der Erstellung von Texten mit einbezogen werden.

Literatur

- Ananny, Mike (2016): Toward an Ethics of Algorithms. In: *Science, Technology, & Human Values* 41 (1), S. 1–25. DOI: 10.1177/0162243915606523.
- boyd, danah; Crawford, Kate (2012): Critical Questions for Big Data. In: *Information, Communication & Society* 15 (5), S. 662–679. DOI: 10.1080/1369118X.2012.678878.
- Brey, Philip (2010): Values in technology and disclosive computer ethics. In: Luciano Floridi (Hg.): *The Cambridge Handbook of Information and Computer Ethics*. Cambridge, MA: Cambridge University Press, S. 41–58.
- Broad, Ellen; Smith, Amanda; Wells, Peter (2017): *Helping organisations navigating ethical concerns in their data practices*. Open Data Institute, <https://de.scribd.com/document/358778144/ODI-Ethical-Data-Handling-2017-09-13>, zuletzt aktualisiert am September 2017 [10.7.2020].
- Burr, Viv (2015): Social Constructionism. In: James D. Wright (Hg.): *International Encyclopedia of the Social & Behavioral Sciences*. Second edition. Amsterdam, Waltham, MA, Oxford: Elsevier Science, S. 222–227.
- Bynum, Terrell Ward (2010): The historical roots of information and computer ethics. In: Luciano Floridi (Hg.): *The Cambridge Handbook of Information and Computer Ethics*. Cambridge, MA: Cambridge University Press, S. 20–38.
- DataEthics.eu (Hg.) (2021): *Data Ethics Readiness Test*. Questionnaire, <https://dataethics.eu/wp-content/uploads/dataethics-readiness-test-2021.pdf> [24.6.2021].

- Dencik, Lina; Hintz, Arne; Redden, Joanna; Treré, Emiliano (2019): Exploring Data Justice: Conceptions, Applications and Directions. In: *Information, Communication & Society* 22 (7), S. 873–881. DOI: 10.1080/1369118X.2019.1606268.
- D'Ignazio, Catherine; Klein, Lauren (2020): *Data Feminism*: MIT Press (Strong Ideas Series).
- Dunkelau, Jannik; Leuschel, Michael (2019): *Fairness-Aware Machine Learning. An Extensive Overview*.
- Fairclough, Norman (1989): *Language and Power*. London: Longman (Language in Social Life Series).
- Fairclough, Norman (1995): *Critical Discourse Analysis. The Critical Study of Language*. London: Longman (Language in Social Life Series).
- Fairclough, Norman (2001): Critical discourse analysis as a method in social scientific research. In: Ruth Wodak und Michael Meyer (Hg.): *Methods of Critical Discourse Analysis*. London: SAGE (Introducing Qualitative Methods), S. 121–138.
- Fairclough, Norman (2010): *Critical Discourse Analysis. The Critical Study of Language*. New York: Routledge.
- Floridi, Luciano (2019): Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. In: *Philos. Technol.* 32 (2), S. 185–193. DOI: 10.1007/s13347-019-00354-x.
- Floridi, Luciano; Taddeo, Mariarosaria (2016): What is data ethics? In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 374 (2083). DOI: 10.1098/rsta.2016.0360.
- Friedman, Batya; Hendry, David G. (2019): *Value Sensitive Design. Shaping Technology with Moral Imagination*. Cambridge, MA: MIT University Press.
- Friedman, Batya; Nissenbaum, Helen (1996): Bias in Computer Systems. In: *ACM Transactions on Information Systems* 14 (3), S. 330–347.
- Ginnis, Steve; Evans, Harry; Boal, Naomi; Davies, Emily; Palmqvist Aslaksen, Alexandra (2016): *Public dialogue on the ethics of data science in government*. Ipsos MORI, <https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf> [29.6.2021].
- Hagendorff, Thilo (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: *Minds and Machines* 30 (1), S. 99–120, <https://arxiv.org/ftp/arxiv/papers/1903/1903.03425.pdf> [29.6.2021].
- Hoffman, Anna Lauren (2019): Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. In: *Information, Communication & Society* 22 (7), S. 900–915. DOI: 10.1080/1369118X.2019.1573912.
- Holland, Martin (2020): "Fuck the algorithm": Proteste in London gegen Corona-bedingte Abi-Notenvergabe, <https://www.heise.de/news/Fuck-the-algorithm-Proteste-in-London-gegen-Corona-bedingte-Abi-Notenvergabe-4872096.html> [6.7.2021].
- Jobin, Anna; Ienca, Marcello; Vayena, Effy (2019): The global landscape of AI ethics guidelines. In: *Nature Machine Intelligence* 2019 (1), S. 389–399. DOI: 10.1038/s42256-019-0088-2.
- Jørgensen, Marianne; Phillips, Louise (2002): *Discourse Analysis As Theory and Method*. London: SAGE.
- La Fors, Karolina; Custers, Bart; Keymolen, Esther (2019): Reassessing values for emerging big data technologies: integrating design-based and application-based approaches. In: *Ethics Inf Technol* 21 (3), S. 209–226. DOI: 10.1007/s10676-019-09503-4.
- Lange, A. R. (2016): *Digital Decisions: Policy Tools in Automated Decision-Making*. Unter Mitarbeit von Jim Dempsey. Hg. v. Center for Democracy & Technology, <https://cdt.org/insights/digital-decisions-policy-tools-in-automated-decision-making/> [29.6.2021].
- Manders-Huits, Noëmi; Zimmer, Michael (2009): Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities. In: *International Review of Information Ethics* 10 (02), S. 37–44.
- Metcalf, Jacob; Moss, Emanuel; boyd, danah (2019): Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. In: *Social Research: An International Quarterly* 82 (2), S. 449–476.
- Mittelstadt, Brent Daniel (2019): Principles alone cannot guarantee ethical AI. In: *Nat Mach Intell* 1 (11), S. 501–507. DOI: 10.1038/s42256-019-0114-4.
- Mittelstadt, Brent Daniel; Allo, Patrick; Taddeo, Mariarosaria; Wachter, Sandra; Floridi, Luciano (2016): The ethics of algorithms: Mapping the debate. In: *Big Data & Society* 3 (2), 1-21. DOI: 10.1177/2053951716679679.
- Moor, James H. (1985): What is Computer Ethics? In: *Metaphilosophy* 16 (4), S. 266–275.
- Ochigame, Rodrigo (2019): *The Invention of "Ethical AI". How Big Tech Manipulates Academia to Avoid Regulation*. The Intercept, <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/> [29.6.2021].
- Osterwalder, Alex; Pigneur, Yves; Bernarda, Greg; Smith, Alan (2014): *Value Proposition Design*. Hoboken, NJ: Wiley (Strategyzer series).
- Saltz, Jeffrey S.; Dewar, Neil (2019): Data science ethical considerations: a systematic literature review and proposed project framework. In: *Ethics Inf Technol* 21 (3), S. 197–208. DOI: 10.1007/s10676-019-09502-5.
- Schäfer, Mirko Tobias; Franzke, Aline (2020): *Data Ethics Decision Aid (DEDA)*. DEDA-edition 3.1, February 2020, https://dataschool.nl/wp-content/uploads/sites/272/2020/04/DEDAWorksheet_ENG.pdf [29.6.2021].
- Schiff, Daniel; Borenstein, Jason; Biddle, Justin; Laas, Kelly (2021): AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection. In: *IEEE Transactions on Technology and Society* 2 (1), S. 31–42.
- Shapiro, Ben Rydal; Meng, Amanda; O'Donnell, Cody; Lou, Charlotte; Zhao, Edwin; Dankwa, Bianca; Hostetler, Andrew (2020): Re-Shape: A Method to Teach Data Ethics for Data Science Education. In: Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn et al. (Hg.): *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20: CHI Conference on Human Factors in Computing Systems. Honolulu HI USA, 25.04.2020–30.04.2020. New York, NY, USA: ACM, S. 1–13.
- Siffels, Lotje; van den Berg, David; Schäfer, Mirko Tobias; Muis, Iris (Forthcoming 2021): *Public Values and Technological Change: Mapping how municipalities grapple with data ethics*. [Under review].
- Simonite, Tom (2021): *What Really Happened When Google Ousted Timnit Gebru*, <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> [29.6.2021].
- Tarrant, Dave; Maddison, James (2021): *The Data Ethics Canvas 2021*. Online verfügbar unter <https://theodi.org/article/the-data-ethics-canvas-2021/>, zuletzt aktualisiert am 28.06.2021, [29.06.2021].
- Trenham, Claire; Steer, Adam (2019): The Good Data Manifesto. In: Angela Daly, S. Kate Devitt und Monique Mann (Hg.): *Good Data*.

Amsterdam: Institute of Network Cultures (Theory on Demand, 29), S. 37–53.

Utrecht Data School (2020): Data Ethics Decision Aid (DEDA). Utrecht University, <https://dataschool.nl/deda/?lang=en> [29.6.2021].

van den Hoven, Jeroen (2010): The use of normative theories in computer ethics. In: Luciano Floridi (Hg.): *The Cambridge Handbook of Information and Computer Ethics*. Cambridge, MA: Cambridge University Press, S. 59–76.

Vryzakis, Alex; Thereaux, Olivier (2020): *How our network is considering data ethics: survey results*. Online verfügbar unter <https://theodi.org/article/how-our-network-is-considering-data-ethics-survey-results/>, zuletzt aktualisiert am 11.03.2020, [29.06.2021].

Whittlestone, Jess; Nyrupe, Rune; Alexandrova, Anna; Cave, Stephen (2019a): The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In: AIES '19 (Hg.): *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. Unter Mitarbeit von Vincent Conitzer, Gillian Hadfield und Shannon Vallor. AIES '19. Honolulu, HI, USA, 27.–28.01.2019. Association for Computing Machinery. New York: Association for Computing Machinery, S. 195–200, <https://dl.acm.org/doi/pdf/10.1145/3306618.3314289> [4.6.2020].

Whittlestone, Jess; Nyrupe, Rune; Alexandrova, Anna; Dihal, Kanta; Cave, Stephen (2019b): *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. Nuffield Foundation; Leverhulme Centre for the Future of Intelligence. London, <https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf> [10.7.2021].



Helena Häußler

Hochschule für Angewandte
Wissenschaften (HAW) Hamburg
Fakultät Design, Medien & Information,
Department Information
Finkenau 35
22081 Hamburg
helena.haeussler@haw-hamburg.de

Helena Häußler ist wissenschaftliche Mitarbeiterin am Department Information an der Hochschule für angewandte Wissenschaften (HAW) Hamburg. Dort promovierte sie, in Kooperation mit der Universität Hildesheim, zum Vertrauen in Suchmaschinen und deren Einfluss auf das Informationsverhalten. Davor studierte sie M. A. Information Science an der Humboldt-Universität zu Berlin und an der University of Copenhagen. Ihren B. A. in Bibliotheksmanagement absolvierte sie an der Fachhochschule Potsdam.