

Schriften zur Medienpädagogik 60

Un|Sichtbarkeiten?

Medienpädagogik, Intersektionalität und Teilhabe

Sabine Eder
Habib Güneşli
Renate Hillen
Claudia Wegener
Rebecca Wienhold (Hrsg.)

Schriften zur Medienpädagogik 60

Dem Bundesministerium für Familie, Senioren, Frauen und Jugend danken wir für die Förderung des vorliegenden Bandes.

Herausgeber

Gesellschaft für Medienpädagogik und Kommunikationskultur in der Bundesrepublik Deutschland (GMK) e. V.

Anschrift

GMK-Geschäftsstelle

Obernstr. 24a

33602 Bielefeld

Fon: 0521/677 88

Fax: 0521/677 29

E-Mail: gmk@medienpaed.de

Homepage: www.gmk-net.de

Für namentlich gekennzeichnete Beiträge sind die Autor*innen verantwortlich.

Redaktion: Sabine Eder, Habib Güneşli, Renate Hillen, Claudia Wegener, Rebecca Wienhold, Tanja Kalwar

Lektorat: Tanja Kalwar

Einbandgestaltung und Titelillustration: Katharina Künkel

© kopaed 2024

Arnulfstr. 205

80634 München

Fon: 089/688 900 98

Fax: 089/689 19 12

E-Mail: info@kopaed.de

Homepage: www.kopaed.de

ISBN 978-3-96848-752-6

Laura Sūna/Dagmar Hoffmann/Anne Mollen

Diskriminierung durch Algorithmen

Überlegungen zur Stärkung KI-bezogener Kompetenzen

Algorithmische Systeme und Systeme Künstlicher Intelligenz (KI) treffen in verschiedensten Lebensbereichen mal mehr mal weniger wichtige Entscheidungen: Sie sortieren Online-Suchergebnisse, kuratieren Inhalte auf Social Media, diagnostizieren (vermeintliche) Krankheiten oder prognostizieren die Arbeitsmarktintegrationschancen von Geflüchteten. Wenn solche automatisierten und selbstlernenden Systeme eingesetzt werden, kann das zu Diskriminierungen führen sowie bestehende Diskriminierungsmuster verstärken. Aus der Forschungsliteratur wissen wir, dass diverse KI-basierte Systeme oftmals Schiefen (Bias) (re-)produzieren und somit Ungleichheit und Exklusion verstärken. Lopez (2021) hat herausgearbeitet, dass bestimmte Gruppen aufgrund eines „racial bias“ unsichtbar, übermäßig sichtbar oder verzerrt abgebildet werden. Als Folge werden strukturelle Ungleichheiten von Algorithmen teils sogar überbetont.

Studien zeigen, dass Nutzer*innen Algorithmen und KI insgesamt ambivalent bewerten. Sie nehmen oft die algorithmische Kuratierung diverser Dienste nicht wahr bzw. schätzen sie nicht als problematisch oder diskriminierend ein (vgl. MeMo:Ki 2021; Overdiek/Petersen 2022; Cousseran et al. 2023). Für ein selbstbestimmtes Leben mit digitalen Medien ist jedoch ein informierter, reflektierter, kritischer und kompetenter Umgang mit KI notwendig (vgl. Digitales Deutschland 2021). Dies trifft auf Gruppen zu, die von (digitaler) Ungleichheit besonders betroffen sind, wie beispielsweise Migrant*innen, Frauen, Menschen im höheren Lebensalter oder mit niedrigem Bildungsniveau. Sie sind Risiken ausgesetzt, die durch KI-kuratierte Informationen verursacht werden, wie z.B. datengesteuerte Manipulation, Verbreitung von Falsch- und Desinformation und auch die Verstärkung von Stereotypen und Diskriminierung (vgl. Neag et al. 2022; Wang et al. 2024).

Ziel des Beitrags ist es aufzuzeigen, wie digitaler Ungleichheit ausgesetzte Gruppen für algorithmen-basierte Diskriminierung sensibilisiert und zu einem kompetenten Umgang ermächtigt werden können. Somit soll Menschen im Sinne gesellschaftlicher Teilhabe ermöglicht werden, Algorithmen für das Erreichen ihrer alltagsweltlichen Ziele so zu nutzen, dass ihnen dabei keine Nachteile entstehen (vgl. Gruber/Hargittai 2023). Zuerst soll erläutert werden, welche Formen und Ursachen es für algorithmen-basierte Diskriminierung bzw. Diskriminierung durch KI-Systeme gibt. Dem schließt sich ein Einblick in Studien zur Wahrnehmung von algorithmen-basierter

Diskriminierung der Nutzer*innen an. Abschließend werden Überlegungen präsentiert, wie ein selbstbestimmter und kompetenter Umgang mit KI gewährleistet werden könnte.

Algorithmen-basierte Diskriminierung – Formen und Ursachen

Technologien der Automatisierung, die landläufig als KI bezeichnet werden, sind allgegenwärtig und entwickeln sich stetig weiter. Ihre verschiedenen Anwendungsbereiche sind ein vieldiskutiertes Thema, gleichwohl kursieren im öffentlichen Diskurs unterschiedliche Begriffe, wenn Automatisierung gemeint ist: u.a. Algorithmen, maschinelles Lernen, regelbasierte KI, generative KI, reaktive KI. Algorithmen beschreiben eine eindeutige Handlungsvorschrift, um ein vorab definiertes Problem zu lösen – und kommen zu diesem Zweck auch im Kontext digitaler Medien zum Einsatz. Abseits von diesem grundlegenden Verständnis werden Algorithmen zunehmend als komplexe Formeln beschrieben, die menschliche und nicht-menschliche Subjekte klassifizieren und hierarchisieren oder Beziehungen zwischen verschiedenen Variablen bestimmen oder vorhersagen. Dabei wird unter dem Begriff „Algorithmus“ häufig mehr als ein bloßer Rechengang verstanden. Hier impliziert man die Umsetzung eines Rechenmodells in ein technologisches Artefakt und schließt gleichzeitig die Konfiguration für die Benutzung durch die Endanwender*innen mit ein (vgl. Heesen/Reinhardt/Schelenz 2021: 133).

Der Begriff KI wird im allgemeinen Sprachgebrauch für sehr unterschiedliche automatisierte Entscheidungssysteme verwendet, zum Beispiel Chatbots, Spracherkennung, Online-Schach oder Sprachübersetzungen. Problematisch an dem Begriff ist, dass nicht nur durch den Begriff, sondern auch durch die anthropomorphisierende Gestaltung von KI-Anwendungen eine menschenähnliche Intelligenz suggeriert wird. KI-Systeme tragen jedoch keine Verantwortung. Sie werden auf Basis von Trainingsdaten angeleitet und sollen aus den Datensätzen selbstständig lernen, d.h. Schlussfolgerungen vornehmen. Trainingsdaten können aus Bildern, Videos, Audio- bzw. Sprachaufnahmen und/oder Texten bestehen. Da Daten von und über Menschen oftmals Vorurteile, Stereotype und Diskriminierungsmerkmale beinhalten, werden diese bei der Mustererkennung von den Systemen mit übernommen und reproduziert.

Diskriminierung durch Algorithmen beschreibt, wie Nutzer*innen aufgrund ihrer persönlichen Daten wie Einkommen, Bildung, Geschlecht, Alter, ethnische Zugehörigkeit oder Religion über automatisierte Entscheidungs- und maschinelle Lernsysteme ungerecht, unethisch oder einfach

nur anders behandelt werden (vgl. Criado/Such 2019: 82). Mit Diskriminierung werden Personen bestimmte Merkmale zugeschrieben, auf deren Grundlage sich wiederum Gruppen bilden, in die Personen eingeordnet werden. Wenn Personen aufgrund dieser Einordnung anders und insbesondere weniger vorteilhaft behandelt werden, wirkt sich dies nachteilig auf Teilhabe, Handlungs- und Selbstbestimmungsmöglichkeiten aus. Algorithmen-basierte Diskriminierung kann, wie jede Form von Diskriminierung, auf der Grundlage von direkten und indirekten Merkmalen erfolgen.

Die Ursachen für algorithmen-basierte Diskriminierung können sowohl (1) in den verwendeten Daten, (2) im Algorithmus selbst als auch (3) in der Art und Weise, wie dieser verwendet wird, liegen. Im Folgenden werden diese drei Punkte näher beleuchtet:

- (1) Diskriminierung auf der Ebene der *verwendeten Trainingsdaten*. Das Phänomen „Sample Bias“ beschreibt, dass die Ergebnisse algorithmischer Entscheidungssysteme durch eine unvollständige oder nicht-repräsentative Zusammensetzung der Trainingsdatensätze verzerrt sind. Ein Beispiel dafür ist die Gesichtserkennungssoftware, die in der Vergangenheit ► **weiße** Männer besser als ► **Schwarze** Frauen erkannt hat. Daten sind in der Regel nicht neutral und spiegeln (Macht-)Verhältnisse der Gesellschaft wider (vgl. AlgorithmWatch.Ch 2023: 4). Das heißt, auch selbst, wenn ein Trainingsdatensatz die soziale Realität gut abbildet, ist er nicht frei von Strukturen sozialer Ungleichheit: Wenn die Input-Daten eines algorithmischen Entscheidungssystems Aussagen über soziale Verhältnisse machen und diese Verhältnisse aber in der Realität von struktureller Ungleichheit geprägt sind, dann werden diese strukturellen Ungleichheiten ebenfalls im Output des Algorithmus reproduziert (vgl. Rentsch 2023: 31). Zudem sind Informationen und Wissen von und über Minderheiten oft seltener in Datensätzen vertreten. Somit sind Daten über manche Gesellschaftsgruppen verfügbar, über andere jedoch nicht.
- (2) Diskriminierung auf der Ebene des *algorithmischen Modells*. Der Kern eines algorithmischen Entscheidungssystems besteht in der Berechnung von Zusammenhängen zwischen Input-Daten und Output-Variablen. Da diese Entscheidungen in der Regel auf Grundlage statistischer Korrelationen getroffen werden, kann es hier durch die Funktionsweise des algorithmischen Entscheidungssystems selbst dann zu diskriminierenden Urteilen kommen, wenn auf den ersten Blick in den verwendeten Daten gar keine Informationen über das diskriminierungsrelevante Merkmal enthalten sind (vgl. ebd.: 32). Hier werden die sogenannten Proxy-Variablen bedeutsam, Variablen die stellvertretend für andere Va-

riablen stehen (zum Beispiel, wenn diese nicht direkt beobachtbar oder messbar sind) (vgl. AlgorithmWatch.Ch 2023: 6). So dürfte ein algorithmisches Entscheidungssystem, das in Bewerbungsprozessen genutzt wird, beispielsweise nicht Personen auf Grundlage ihres Alters ablehnen, da Alter nach Allgemeinem Gleichbehandlungsgesetz (AGG) ein geschütztes Merkmal ist. Als Proxy-Variable könnte das System aber die Dauer der bisherigen Berufserfahrung der Person nutzen, um dennoch ältere Menschen zu identifizieren und aus dem Bewerbungsprozess auszuschließen (vgl. Digital Autonomy Hub 2022: 8). Zudem können in der Modellarchitektur, und nicht nur in den genutzten Datensätzen, diskriminierende Outputs der Systeme angelegt sein.

- (3) Diskriminierung auf der Ebene der *Einbettung und Anwendung des algorithmischen Systems im gesellschaftlichen Kontext*. Algorithmische Systeme liefern Ergebnisse, die zu Lösungen und Entscheidungen beitragen sollen. Inwiefern sie das tun, hängt unter anderem davon ab, wie Anwender*innen und Entscheidungsträger*innen der Systeme die Ergebnisse interpretieren und ggf. hinterfragen. Rentsch (2023: 33) betont, dass trotz der Intransparenz und fehlenden Erklärbarkeit von KI-Systemen algorithmisch berechnete Entscheidungen grundsätzlich ein großes gesellschaftliches Vertrauen genießen. Menschen tendieren dazu, algorithmen-basierte Systeme in ihrer Leistungsfähigkeit zu überschätzen und sich auf Empfehlungen durch diese zu verlassen. Als Folge dessen passiert es, dass Menschen die Outputs von Algorithmen diskriminierend einsetzen, obwohl die Algorithmen selbst nicht diskriminierend wirken. Diese Tendenz wird als „Automation Bias“ bezeichnet. Dafür benötigen die Beteiligten Kompetenzen, um die Ergebnisse des Systems einschätzen und bewerten sowie die Empfehlungen und Prognosen angemessen hinterfragen zu können (vgl. AlgorithmWatch.Ch 2023: 5).

Einzelne Faktoren der algorithmen-basierten Diskriminierung stehen miteinander in Wechselwirkung und lassen sich nicht immer trennscharf nur einer der Ebenen zuordnen (vgl. Rentsch 2023: 30-31). Diskriminierung kann oft auch aus der Kombination unterschiedlicher geschützter Merkmale entstehen, beispielsweise, wenn Menschen gleichzeitig aufgrund ihrer Geschlechtsidentität und ihrer ethnischen oder sozialen Herkunft diskriminiert werden. Zudem können sogenannte „lernende Systeme“, die selbst Muster identifizieren, neue Merkmale mit diskriminierenden Folgen hervorbringen (vgl. AlgorithmWatch.Ch 2023: 6). Des Weiteren betont die Nichtregierungsorganisation AlgorithmWatch, dass Diskriminierungen in

den verschiedensten Anwendungsbereichen stattfinden können und alle Gruppen betreffen – jene, die durch das bestehende Diskriminierungsverbot geschützt sind, sowie andere, die dies nicht sind, weil das Merkmal, aufgrund dessen die Mitglieder der Gruppe diskriminiert werden, nicht rechtlich aufgeführt ist (z.B. Gewicht, Rechtsstatus) (vgl. ebd.; Wachter 2022). Besonders anfällig für algorithmen-basierte Diskriminierung sind solche Bereiche wie der Zugang zu Sozialleistungen, der Bildungsbereich, der Arbeits- sowie Gesundheitsbereich, die Justiz und die vorausschauende Polizeiarbeit/predictive policing (vgl. AlgorithmWatch.Ch 2023).

Wahrnehmung von algorithmen-basierter Diskriminierung

Diskriminierung durch ein algorithmisches System ist oft sowohl für Betroffene als auch für andere Anwender*innen nicht erkennbar oder wird erst registriert, wenn es zu einer Irritation kommt oder etwas beispielsweise in der Ausgabe oder Anzeige auf Social Media schiefeht. Manche Autor*innen sprechen von „opferloser Diskriminierung“, da die Betroffenen in der Regel gar nicht erfahren, dass sie von Diskriminierung betroffen waren oder sind (vgl. Spiecker 2023). Die Diskriminierung kann im System selbst verankert sein und damit eine große Anzahl von Menschen betreffen (Skalierungseffekt). Dies macht es für einzelne Nutzer*innen schwieriger, die Diskriminierung zu erkennen und sich als davon betroffen zu identifizieren (vgl. AlgorithmWatch.Ch 2023: 6). Sie haben nicht ohne Weiteres Zugang zu allen Daten, die Tech-Unternehmen über sie gesammelt, und dem Wissen, das sie über die eigene Person angeeignet haben. Nutzer*innen können zwar nach der DSGVO von ihrem Auskunftsrecht Gebrauch machen und auf die Herausgabe einer Daten-Kopie pochen. Doch oftmals werden die Daten so übermittelt, dass sie aufgrund der Form, Sprache und Art der Darstellung schwer verständlich sind. Auch lässt sich kaum erkennen, auf welcher Grundlage welche Entscheidungen von dem System getroffen worden sind: ob ein Vertrag zu diesen oder anderen Bedingungen angeboten oder ganz verweigert wird. Nutzer*innen können eher selten rückschließen, warum ihnen welche Lösungen angeboten werden oder auch nicht.

In einer qualitativen Studie zu Diskriminierungserfahrungen von Menschen mit Migrationsgeschichte, bei der 25 Personen in Gruppendiskussionen befragt wurden, zeigte sich (vgl. Süna/Hoffmann 2024), dass einige Teilnehmer*innen den Umgang mit algorithmischen Medien als störend, unangenehm oder zumindest irritierend empfanden. Als unerwünscht wurde die algorithmische Klassifizierung auf der Grundlage der kulturellen Zugehörigkeit beschrieben sowie die automatische Zuordnung der

Nutzer*innen zu einem Mehrheitssprachensystem (Wörter und Namen werden ungefragt angepasst) sowie die Nivellierung und Standardisierung auf der Grundlage von Kriterien, die vom KI-System festgelegt wurden. Zuweilen sind die befragten Nutzer*innen aber auch zufrieden damit, wie sie von Empfehlungsalgorithmen klassifiziert werden, empfinden sie als bequem und praktisch (z.B. wenn ihnen bei den Streamingdiensten bestimmte Filmvorschläge gemacht werden). Obgleich eine gelegentliche Überbewertung ihrer Interessen durch Algorithmen Irritationen hervorruft, haben nur wenige der Befragten der fünf Fokusgruppen von explizit diskriminierenden Auswirkungen auf die Nutzer*innenprofilierung berichtet. Algorithmen-basierte Zuschreibungen einer bestimmten kulturellen Zugehörigkeit in Form von Empfehlungsalgorithmen von Streaming- oder Dating-Apps sowie das Nicht-dazu-gehören, was durch die KI-basierte Autokorrektur ausländisch klingender Namen suggeriert wird, wird eher relativiert.

Nur wenige der Teilnehmer*innen der genannten Studie von Süna und Hoffmann (2024) fühlen sich motiviert und in der Lage, aktiv auf KI-Technologien einzuwirken, um eine Ungleichbehandlung zu minimieren resp. zu verhindern. Bei den meisten Nutzer*innen hat sich eine Art digitale Resignation eingestellt (vgl. auch Draper/Turow 2019), sie sehen kaum Interventionsmöglichkeiten. Ferner verfügen sie über begrenzte Fähigkeiten, Einstellungen zu verändern oder Umprogrammierungen vorzunehmen, um KI-Technologie zweckmäßiger und diskriminierungsfreier zu gestalten. In den Fokusgruppen wurde deutlich, dass es an einer Sensibilisierung für Diskriminierungspotenziale mangelt, wobei Menschen mit Migrationsgeschichte ein besonders großes Risiko haben, Ungleichheiten und Ausgrenzungen bei bestimmten digitalen Anwendungen zu erfahren. Grund dafür könnte sein, dass es an vergleichenden Negativbeispielen bei der Nutzung von KI-Systemen fehlt, zu denen sie ihre eigene Nutzungsweise und zugehörige Erfahrungen in Beziehung setzen könnten (auch im Sinne geteilter Betroffenheit). Zudem werden diskriminierende Potenziale von KI-Technologien im öffentlichen Diskurs kaum angesprochen (z.B. das Online-Profiling und Targeting in der öffentlichen Verwaltung, in der Privatwirtschaft oder auf Online-Plattformen) (vgl. Süna/Hoffmann 2024), bzw. es wird darüber noch zu wenig aufgeklärt.

Aus den wenigen vorliegenden Studien wird deutlich, dass es in der Wahrnehmung algorithmen-basierter Diskriminierung soziodemografische Unterschiede gibt und das unter anderem Jugendliche dafür sensibilisierter sind als ältere Nutzer*innengruppen. So zeigt die Studie von Schober und Kolleg*innen (2022), dass Jugendliche über Diskriminierungsrisiken auf Plattformen mit Algorithmischen Empfehlungssystemen (AES) informiert

sind und sie Diskriminierung dabei auf zwei Ebenen wahrnehmen: Einerseits würden – so die Jugendlichen – durch Algorithmen diskriminierende Inhalte wie ► **LGQTIA***-feindliche Videos oder Inhalte, in denen bestimmte Nationalitäten, Religionen oder Menschen mit Behinderung beleidigt werden, verstärkt zirkulieren. Somit würden solche Inhalte mehr Aufmerksamkeit erlangen. Andererseits würden Algorithmen Inhalte mancher Gesellschaftsgruppen besonders stark verbreiten und somit andere Gruppen ausschließen. Beispielsweise werden nicht-migrantische, traditionell agierende Menschen der Mehrheitsgesellschaft und deren Werte sowie Schönheitsideale durch AES bevorzugt. Darauf verweist auch Bozdağ (2022) in ihrer Untersuchung und konnte zeigen, dass migrantische Jugendliche in Online-Umgebungen Ungleichbehandlung aufgrund ihrer Herkunft erleben (siehe auch den Beitrag von Bozdağ in diesem Band).

KI-bezogene Kompetenzen

Um eine Sensibilisierung für mögliche Diskriminierung zu erreichen und Diskriminierungsrisiken zu mindern (vgl. u.a. Knaus 2024), plädieren wir für eine stärkere Unterstützung der KI-bezogenen Medienkompetenzaneignung. Ausgehend von vorliegenden und eigenen Studien zu digitalen Medienkompetenzen im Kontext von KI, umfassen KI-bezogene Kompetenzen folgende Aspekte (vgl. Brüggem/Süna 2023; Dogruel 2021; Digitales Deutschland 2021):

- das *Bewusstsein* und die *Wahrnehmung* der Allgegenwärtigkeit von algorithmen-basierten und KI-getriebenen Prozessen in digitalen Medien und Anwendungen (Online-Apps, Social-Media-Plattformen, Online-Dienste, Chat-Bots);
- *allgemeines Wissen* über die Funktionsweise von KI-Anwendungen;
- die Fähigkeit, KI-geprägte Entscheidungen *kritisch* und *affektiv zu bewerten* und zu hinterfragen und die Konsequenzen für sich und die Gesellschaft einschätzen zu können;
- die Fähigkeit, mit der Prägekraft von KI-Anwendungen umzugehen sowie diese ggf. *kreativ zu bewältigen* und zu *beeinflussen*, sowie KI-Anwendungen sozial verantwortlich *zu nutzen*.

Insgesamt werden hier drei Kompetenzdimensionen angesprochen: die kognitive, die affektive und die handlungsbezogene Dimension (siehe Abb. 1). Der Umgang mit KI-Technologien betrifft zum Beispiel im kognitiven Bereich die Wahrnehmung und Bedeutungskonstruktion über KI sowie im affektiven Bereich die emotionale und affektive Einordnung des Umgangs

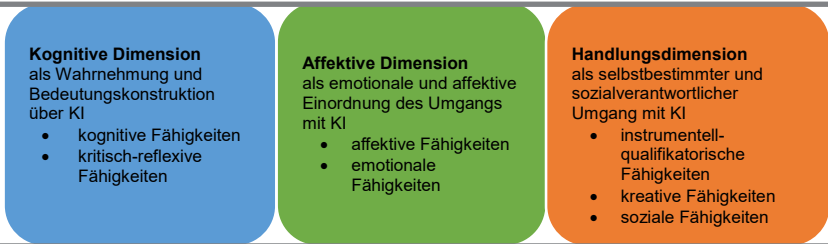


Abb. 1: Dimensionen von KI-bezogenen Kompetenzen (Quelle: Eigene Darstellung)

mit KI. Die Handlungsdimension umfasst dementsprechend Praktiken des kreativen, selbstbestimmten und sozialverantwortlichen Umgangs mit KI-Technologien. Eine KI-kompetente Person ist in der Lage, fundierte Bewertungen über KI-Technologien vorzunehmen und mit den ethischen Fragen umzugehen, die sich dabei stellen.

Bezogen auf die beschriebenen Ursachen algorithmen-basierter Diskriminierung bedeutet dies, dass kompetente Nutzer*innen zumindest über allgemeines Wissen um die Fehlbarkeit von Algorithmen verfügen müssen. Sie können kritisch das Zustandekommen und die Funktion von algorithmischen Modellen in digitalen Medien einschätzen sowie die gesellschaftlichen Folgen solcher Modelle kritisch-reflexiv bewerten. Da algorithmen-basierte Diskriminierung oft durch irritierende Nutzungserfahrung sichtbar wird, ist somit die Fähigkeit, das Nutzungserlebnis affektiv und emotional einzuordnen, notwendig. Das zeigt sich beispielsweise, indem man die Algorithmen als Ursache der Irritation wahrnehmen und die damit verbundenen Emotionen des Genervt-Seins oder der Überraschung auf die Algorithmen zurückführen kann. Im Weiteren wollen wir einige Überlegungen zur Stärkung von KI-bezogenen Kompetenzen aufzeigen.

Überlegungen zur Stärkung KI-bezogener Kompetenzen: Informieren, sensibilisieren, ermächtigen

Medienpädagogische Arbeit kann mit dem Ziel der Stärkung von KI-bezogenen Kompetenzen auf drei Ebenen ausgerichtet sein: über Algorithmen und KI-Systeme *informieren*, über die Gefahren möglicher Diskriminierung *sensibilisieren* und den Nutzenden *Strategien* des selbstbestimmten *Umgangs* an die Hand geben.

Folgende, an Kalogeropoulos et al. (2021) angelehnte, Übersicht beschreibt die Besonderheiten algorithmen-basierter Diskriminierung, die für die

Wahrnehmung von KI-Systemen in digitalen Mediumgebungen bedeutsam sein können (siehe Tab. 1). Dabei werden vorwiegend kognitive sowie kritisch-reflexive Fähigkeiten angesprochen, zudem gewinnt auch die affektive Dimension an Bedeutung. An diese Kompetenzdimensionen können mögliche Handlungspraktiken anschließen, die weiter unten erläutert werden.

Besonderheiten algorithmen-basierter Diskriminierung	Ausdrucksformen	Kompetenz- dimension
Versteckt	Algorithmische Systeme werden (oft) in nicht öffentlich zugänglichen Prozessen eingesetzt und über ihren Einsatz wird nicht hinreichend informiert	Kognitive Fähigkeiten
Opak	Algorithmische Systeme sind nicht immer automatisch für Menschen einsehbar und verständlich	Kognitive Fähigkeiten
Indirekt	Algorithmische Systeme ermöglichen Diskriminierung auf Basis von Ersatzidentitäten und nicht geschützten Diskriminierungsmerkmalen (Proxy-Variablen)	Kritisch-reflexive Fähigkeiten
Systemisch	Algorithmen-basierte Diskriminierung wird teils erst im Gruppenbezug erkennbar	Kritisch-reflexive und affektive Fähigkeiten
Intersektional	Algorithmen-basierte Diskriminierung kann gegen eine Schnittmenge verschiedener Gruppen gerichtet sein	Kritisch-reflexive Fähigkeiten
Scheinbar objektiv	Entscheidungen algorithmischer Systeme „wirken“ objektiv	Affektive Fähigkeiten
Expliziert	Algorithmen-basierte Diskriminierung spiegelt bestehende, teils versteckte Diskriminierungsstrukturen wider. Letztere werden damit technisch expliziert und für eine öffentliche Debatte zugänglich	Kritisch-reflexive Fähigkeiten

Tab. 1. Besonderheiten algorithmen-basierter Diskriminierung und entsprechender Medienkompetenzdimension (Quelle: Eigene Darstellung in Anlehnung an Kalogeropoulos et al. 2021: 9)

Weitere Aspekte und Fähigkeiten für die Sensibilisierung über mögliche algorithmen-basierte Diskriminierung sind (ohne einen Anspruch auf Vollständigkeit):

- Die Beschäftigung mit algorithmischen Systemen, ihren Fallstricken und Besonderheiten, hilft nicht nur verstehen, in welchen Bereichen diese zum Einsatz kommen (können), sondern sensibilisiert darüber hinaus für Folgewirkungen, die bei ihrer Anwendung auftreten können (vgl. Kalogeropoulos et al. 2021: 4).
- Die Rolle von Menschen bei der Entwicklung von KI verstehen und erkennen können. Erkennen, dass Menschen eine wichtige Rolle bei der Programmierung, der Auswahl von Modellen und der Optimierung von KI-Systemen spielen. Dies hilft, die Technologie nicht als objektiv zu bewerten, sondern als fehleranfällig (vgl. Long/Magerko 2020: 4-5).
- Stärken und Schwächen von KI einschätzen können. Situationen erkennen, bei denen KI an ihre Grenzen stößt und die für KI-Systeme eine größere Herausforderung darstellen. Im Kontext dieser Informationen entscheiden, wann der Einsatz von KI sinnvoll ist und wann menschliche Fähigkeiten genutzt werden sollten (vgl. ebd.).
- Unterschiedliche Perspektiven auf die wichtigsten ethischen Fragen im Zusammenhang mit KI identifizieren und beschreiben können, insbesondere im Kontext von Datenschutz, Recruiting, Fehlinformationen, ethischer Entscheidungsfindung, Voreingenommenheit, Transparenz und Verantwortung (vgl. ebd.).
- Ein kritisches Verständnis über die Beziehung zwischen privaten Daten und algorithmischem Output entwickeln können (vgl. Gruber/Hargittai 2023: 11).

Welche Handlungsstrategien können nun Nutzer*innen, aufbauend auf ihrem Wissen über Algorithmen und KI-Systeme, in ihrem alltäglichen Medienhandeln umsetzen, damit algorithmen-basierte Diskriminierung adressiert werden kann? Auch wenn allgemein die Affordanzen sowie Infrastrukturen der meisten Plattformen die Gestaltungsmöglichkeiten für ‚einfache‘ Nutzer*innen prinzipiell doch eher stark eingrenzen und es den Anschein haben mag, dass die Nutzer*innen einen Teil ihrer Handlungsfähigkeit einbüßen, sobald sie in algorithmische Prozesse involviert sind, so gibt es doch auch Möglichkeiten der aktiven Beteiligung und Steuerung im Umgang mit KI-Technologien. KI-Anwendungen können und werden zuweilen von Nutzer*innen eigensinnig, kreativ und produktiv genutzt. Unter anderem lassen sich folgende Fähigkeiten zum kompetenten Umgang mit Algorithmen und KI-Systemen ausmachen bzw. sind wünschenswert:

- Kreativer Umgang mit KI-Technologien als Praktiken des Austricks des Algorithmus. Das sind auf alltägliches Nutzungswissen basierende Bemühungen, die Online-Ausgabe so zu beeinflussen, dass Nutzer*innen ihre Ziele erreichen können. Beispielsweise, wenn man bei der Online-Suche unterschiedliche Webbrowser verwendet oder bei einer Dating-App aktiv versucht, die Anzeige zu verändern, indem man bewusst unpassende Vorschläge ablehnt und wegeklickt. Studien zeigen, dass einige, insbesondere jüngere Befragte diese Praktik ganz selbstverständlich anwenden (vgl. Sūna/Hoffmann 2024).
- Gezielte Nutzung der plattformeigenen Gestaltungsmöglichkeiten im Kontext von Datenschutz. Dies kann sich von bewusster Begrenzung der Zugriffsrechte, beispielsweise bei Facebook oder Instagram, bis hin zur Nutzung des Inkognito-Modus in einem Web-Browser erstrecken.
- Profundes Wissen über KI ermöglicht es Nutzer*innen, Strategien anzuwenden, die ihnen mehr Kontrolle darüber geben, welche persönlichen Daten gesammelt und möglicherweise für die Anpassung von Inhalten notwendig werden (vgl. Gruber/Hargittai 2023: 11-12).
- Praktische Einblicke in die Programmierung algorithmischer Systeme fördern das Verständnis über deren Möglichkeiten und Grenzen.
- Nutzer*innen muss die Gelegenheit gegeben werden, diskriminierende Erfahrungen in KI-basierten Medienumgebungen publik machen zu können. Hier können Austauschforen (zum Beispiel über Apps) hilfreich sein, um systematische Diskriminierungen durch algorithmische Entscheidungssysteme zu identifizieren und ihnen zu begegnen (vgl. Heesen/Reinhard/Schelenz 2021).
- Nutzer*innen sollten sich ermutigt fühlen, ihre Verbraucherrechte geltend zu machen und Meldesysteme zu frequentieren, wenn sie sich benachteiligt und angegriffen fühlen.

Ausblick

Die Ausführungen haben klar gemacht, dass es sich bei algorithmen-basierter Diskriminierung um ein vielschichtiges und komplexes Phänomen handelt, das sowohl bestehende Medienkompetenzansätze als auch Antidiskriminierungsgesetzgebung herausfordert. Die Politik hat wiederholt, beispielsweise durch die Datenethikkommission der Bundesregierung¹, bescheinigt bekommen, dass das Allgemeine Gleichbehandlungsgesetz (AGG) in Deutschland nicht ausreichend vor Diskriminierung durch Algorithmen schützt. Diese Lücke besteht, obwohl Diskriminierung durch Algorithmen durch die technische Skalierung potenziell massenhaft erfolgen kann und

zugleich oft im Verborgenen stattfindet, ohne dass sich die Betroffenen überhaupt darüber bewusst werden. Genau aus diesem Grund braucht es ein Ineinandergreifen von regulatorischer Anpassung und Weiterentwicklung von Medienkompetenzansätzen. Denn es wäre falsch, die Last gegen algorithmen-basierte Diskriminierung vorzugehen, alleinig auf die Schultern der Betroffenen zu laden. Diese Menschen sind in der Regel bereits besonders schutzbedürftig, so dass es eines starken regulatorischen Rahmens bedarf. Dennoch kann auch ein erfolgreicher Schutz vor Diskriminierung im Rahmen des AGG nur erfolgen, wenn Menschen für algorithmen-basierte Diskriminierung sensibilisiert sind. Mit anderen Worten, sie müssen über entsprechende Kompetenzen verfügen, um überhaupt bei einer Andersbehandlung an Algorithmen als mögliche Verursacher zu denken und diese beispielsweise gezielt versuchen zu überprüfen. Der von offiziellen Stellen geforderte Ausbau von Beratungsstrukturen, um besser gegen algorithmen-basierte Diskriminierung vorzugehen, kann daher nur funktionieren, wenn Menschen mit einem Anfangsverdacht diese Beratungsstellen aufsuchen. Für diese Kompetenzentwicklung braucht es auch den politischen Willen, eine umfassende medienpädagogische Arbeit zu fördern und zu finanzieren, die sich nicht nur auf den Kinder- und Jugendbereich fokussiert. Denn Menschen in allen Lebensbereichen können von algorithmen-basierter Diskriminierung betroffen sein. Durch die entsprechende Sensibilisierung können auch zivilgesellschaftliche Organisationen ihre Unterstützungsangebote deutlich besser koordinieren und ein möglicherweise zu etablierendes Verbandsklagerecht erfolgreich nutzen. Hierfür braucht es aber Präzedenzfälle aus der alltäglichen Lebenswelt von Betroffenen – deren kritisches Reflexionsvermögen Dreh- und Angelpunkt für ein zielgerichtetes Vorgehen gegen algorithmen-basierte Diskriminierung bleibt.

Anmerkung

- 1 www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html [Stand: 01.07.2024]

Literatur

AlgorithmWatch.Ch (2023): Positionspapier – Schutz vor algorithmischer Diskriminierung. Abrufbar unter: <https://algorithmwatch.ch/de/diskriminierende-algorithmen/> [Stand: 01.07.2024].

- Bozdağ, Çiğdem (2022): Inclusive Media Education in the Diverse Classroom: A Participatory Action Research in Germany. In: *Media and Communication*, 10(4), 305-316.
- Cousseran, Laura/Lauber, Achim/Herrmann, Simon/Brüggen, Niels (2023): *Kompass: Künstliche Intelligenz und Kompetenz 2023. Einstellungen, Handeln und Kompetenzentwicklung im Kontext von KI*. München: kopaed.
- Criado, Natalia/Such, Jose M. (2019): Digital Discrimination. In: Karen Yeung/Martin Lodge (Hrsg.): *Algorithmic Regulation*. Oxford University Press, 82-85.
- Digital Autonomy Hub (2022): *Algorithmenbasierte Diskriminierung. Warum Antidiskriminierungsgesetze jetzt angepasst werden müssen*. POLICY BRIEF # 5. Abrufbar unter: https://algorithmwatch.org/de/wp-content/uploads/2022/02/DAH_Policy_Brief_5.pdf [Stand: 01.07.2024].
- Digitales Deutschland (2021): *Digitales Deutschland. Ein Rahmenkonzept*. Abrufbar unter: https://digid.jff.de/wp-content/uploads/2021/06/Rahmenkonzept_Digitales_Deutschland_Vollversion.pdf [Stand: 01.07.2024].
- Dogruel, Leyla (2021): What is Algorithm Literacy?: A conceptualization and challenges regarding its empirical measurement. In Taddicken, M./ Schumann, C. (Hrsg.): *Algorithms and Communication*. Freie Universität Berlin, 67-93.
- Draper, Nora A./Turow, Joseph (2019): The corporate cultivation of digital resignation. In: *New Media & Society*, 21(8), 1824-1839.
- Gruber, Jonathan/Hargittai, Eszter (2023): The importance of algorithm skills for informed Internet use. In: *Big Data & Society*, 10(1), 1-14.
- Heesen, Jessica/Reinhardt, Karoline/Schelenz, Laura (2021): *Diskriminierung durch Algorithmen vermeiden. Analysen und Instrumente für eine digitale demokratische Gesellschaft*. In: Bauer, Gero/Engelmann, Sebastian/Haug, Lean/Kechaja, Maria (Hrsg.): *Diskriminierung/Antidiskriminierung*. Bielefeld: transcript, 129-148.
- Kalogeropoulos, Elena/Lammers, Anne/Müller-Brehm, Jaana/Puntschuh, Michael (2021): *Wegweiser Digitale Debatten. Teil 2: Algorithmenvermittelte Diskriminierung*. Abrufbar unter: www.bmfsfj.de/resource/blob/186300/961021829a491933cf24e8f06ff8018f/wegweiser-digitale-debatten-teil-2-data.pdf [Stand: 01.07.2024].
- Knaus, Thomas (2024): Warum KI kein Hype ist und die Medienpädagogik sich damit beschäftigen sollte. In: *medien + erziehung (merz)*, 68. Jg., H. 3, 21-30.
- Long, Duri/Magerko, Brian (2020): What is AI literacy? Competencies and design considerations. In: *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: ACM, 1-16.
- Lopez, Paola (2021): *Diskriminierung durch Data Bias. Künstliche Intelligenz kann soziale Ungleichheiten verstärken*. In: *WZB Mitteilungen*, Heft 171, 26-28.

- MeMo:KI (2020): Meinungsmonitor Künstliche Intelligenz. Künstliche Intelligenz und Diskriminierung. Factsheet Nr. 2. Abrufbar unter: www.cais-research.de/wp-content/uploads/Factsheet-2-Diskriminierung.pdf [Stand: 01.07.2024].
- Neag, Annamária/Bozdağ, Çiğdem/Leurs, Koen (2022): Media Literacy Education for Diverse Societies. In: Neag, A./Bozdağ, Ç./Leurs, K.: Oxford Research Encyclopedia of Communication. Oxford University Press, 1-22.
- Overdiek, Markus/Petersen, Thomas (2022): Was Deutschland über Algorithmen und Künstliche Intelligenz weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage: Update 2022. Bertelsmann Stiftung. Abrufbar unter: www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/DG_Was_Deutschland_ueber_Algorithmen_KI.pdf [Stand: 01.07.2024].
- Rentsch, Susanne (2023) „Computer sagt nein“ – Gesellschaftliche Teilhabe und strukturelle Diskriminierung im Zeitalter Künstlicher Intelligenz. In: Wager, Andreas/Stark, Carsten (Hrsg.): Die Digitalisierung des Politischen, Sozialwissenschaften und Berufspraxis. Wiesbaden: Springer VS, 23-44.
- Schober, Maximilian/Lauber, Achim/Bruch, Louisa/Herrmann, Simon/Brüggen, Niels (2022): „Was ich like, kommt zu mir.“ Kompetenzen von Jugendlichen im Umgang mit algorithmischen Empfehlungssystemen. München: kopaed.
- Spiecker, gen. Döhmman, Indra/Towfigh, Emanuel V. (2023): Automatisch benachteiligt. Das Allgemeine Gleichbehandlungsgesetz und der Schutz vor Diskriminierung durch algorithmische Entscheidungssysteme. Rechtsgutachten im Auftrag der Antidiskriminierungsstelle des Bundes. Abrufbar unter: www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Rechtsgutachten/schutz_vor_diskriminierung_durch_KI.pdf?__blob=publicationFile&v=9 [Stand: 01.07.2024].
- Süna, Laura/Hoffmann, Dagmar (2024): From AI imaginaries to AI literacy: Artificial intelligence technologies in everyday lives of migrants in Germany. In: *MedieKultur*, Vol. 40, No. 76 (2024): Special Issue: AI and communicative practices, 53-76.
- Wachter, Sandra (2022): The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law. In: *SSRN Electronic Journal*, 97 (1), 149-204.
- Wang, Chenyue/Boerman, Sophie C./Kroon, Anne C./Möller, Judith/de Vreese, Claes H. (2024): The artificial intelligence divide: Who is the most vulnerable? In: *New Media & Society*, 1-23.

Förderhinweis

Die Publikation ist im Rahmen des Projekts Digitales Deutschland | Monitoring zur Digitalkompetenz der Bevölkerung entstanden. Das Projekt wird vom Bundesministerium für Familie, Senioren, Frauen und Jugend gefördert und vom JFF – Institut für Medienpädagogik in Forschung und Praxis in Kooperation mit der Universität Siegen und der PH Ludwigsburg durchgeführt.

Lizenz

Der Artikel steht unter der Creative Commons Lizenz **CC BY-SA 4.0**. Die Namen der Urheberinnen sollen bei einer Weiterverwendung genannt werden. Wird das Material mit anderen Materialien zu etwas Neuem verbunden oder verschmolzen, sodass das ursprüngliche Material nicht mehr als solches erkennbar ist und die unterschiedlichen Materialien nicht mehr voneinander zu trennen sind, muss die bearbeitete Fassung bzw. das neue Werk unter derselben Lizenz wie das Original stehen. Details zur Lizenz: <https://creativecommons.org/licenses/by-sa/4.0/legalcode>.

Einzelbeiträge werden unter www.gmk-net.de/publikationen/artikel veröffentlicht.

Glossar

Abled bodied

Kann mit „leistungs- oder arbeitsfähig“ übersetzt werden. Able-bodied bezeichnet die gesellschaftlich privilegierte Position von Menschen, die keine Behinderung haben.

Agender

Mit dem Begriff bezeichnen sich Menschen, die kein Geschlecht haben, sich keinem Geschlecht zugehörig fühlen oder mit dem Konzept von Geschlecht nichts anfangen können.

Be_hindert

Der Unterstrich wird häufig im aktivistischen Bereich eingesetzt, um zu zeigen: Behindert ist man nicht – behindert wird man. Er soll zeigen, dass Behinderung durch äußere Umstände und Barrieren im Alltag produziert wird und nicht der Körper einer Person das Problem ist.

BIPoC/PoC

BIPOC steht für Black, Indigenous, People of Color (Schwarz, indigen, People of Color). Diese Abkürzung wird oftmals verwendet, um die Diskriminierungserfahrungen von Schwarzen Menschen als auch indigenen Gruppen besonders hervorzuheben.

Cis (gender)

„Cis“ ist das Gegenstück zu „trans“. „Cis“ wird benutzt, um auszudrücken, dass eine Person das Geschlecht hat, dem sie bei der Geburt aufgrund der Genitalien zugewiesen wurde und sich entsprechend identifiziert. Als Beispiel: Eine cis Frau ist eine Person, die bei der Geburt dem weiblichen Geschlecht zugewiesen wurde und sich auch als Frau identifiziert.

Cisnormativität

Cisnormativität ist ein Teil von Heteronormativität. Es wird davon ausgegangen, dass alle Menschen cisgeschlechtlich sind, womit trans Personen abgewertet und unsichtbar gemacht werden.

Endogeschlechtlich

Der Begriff „endogeschlechtlich“ oder „endo“ (griech. „éndon“: innen, innerhalb) beschreibt Menschen, die nicht inter* sind, das heißt, deren Kör-

per sich nach medizinischen Normen vermeintlich eindeutig als nur weiblich oder nur männlich einordnen lassen.

FLINTA/MINTA

FLINTA steht für Frauen, Lesben, inter, nicht-binäre, trans und agender Personen. Bei der Abkürzung MINTA wird das F und das L mit dem M für Mädchen getauscht, um gezielt ein jüngerer Publikum anzusprechen.

Fremdouting

Bezeichnet die (absichtliche oder unabsichtliche) unfreiwillige Offenlegung der sexuellen Orientierung oder Geschlechtsidentität einer Person durch eine dritte Person. Ein bekanntes Beispiel ist das Fremdouting der beiden Prominenten Hape Kerkeling und Alfred Biolek im Jahre 1991.

Gender

Der englische Begriff „gender“ bezeichnet das durch Gesellschaft und Kultur geprägte soziale Geschlecht in Abgrenzung zum biologischen Geschlecht. Gemeint sind damit Erwartungen, Rollen und Werte, die an das bei der Geburt zugewiesene Geschlecht geknüpft sind.

Heteronormativität

Der Begriff bezieht sich auf die Annahme, dass es grundsätzlich nur zwei Geschlechter (weiblich und männlich) gäbe, die sich gegenseitig sexuell begehren. Menschen, die nicht in die zweigeschlechtliche Ordnung passen, weil sie sich zum Beispiel als trans* identifizieren, werden als „anders“ wahrgenommen, weil sie von der Norm abweichen.

LGBTQIA*

Steht für Lesbian (Lesbisch), Gay (Schwul), Bisexual (Bisexuell), Transgender, Queer, Intersex (Intergeschlechtlich) und Asexual (Asexuell) und umfasst verschiedene sexuelle Orientierungen und geschlechtliche Identitäten.

Misgendern

Bedeutet, über eine Person mit falschen Pronomen oder falsch gegenderten Begriffen zu sprechen oder eine falsche Anrede zu verwenden, die nicht dem Geschlecht der Person entspricht. Das passiert zum Beispiel, wenn eine nichtbinäre Person mit „Frau“ angeredet wird. Personen können absichtlich oder unabsichtlich eine Person misgendern.

Nonbinär

Als nonbinär können sich Menschen bezeichnen, die nicht (oder nicht zu 100%) Mann oder Frau sind. Stattdessen ist ihr Geschlecht beispielsweise beides gleichzeitig, zwischen männlich und weiblich, oder weder männlich noch weiblich. Manche nichtbinäre Menschen verorten sich ganz außerhalb des binären Systems, manche haben gar kein Geschlecht (agender) oder haben eine Geschlechtsidentität, die sich immer wieder ändert (genderfluid).

Queer

Der Begriff wird zum einen als Überbegriff für Menschen verwendet, die nicht in die geschlechtliche und/oder sexuelle Norm (hetero) passen. Zudem wird der Begriff auch als Selbstbezeichnung von Menschen verwendet, um eine Offenheit für die sexuelle und geschlechtliche Vielfalt zu zeigen.

Schwarz/weiß

Schwarz ist eine Selbstbezeichnung, die Menschen mit afrikanischer, karibischer und afroamerikanischer Herkunft verwenden. Es wird großgeschrieben, weil es dabei nicht um die Hautpigmentierung geht, sondern um ein soziales und politisches Konstrukt. Auch beim Begriff *weiß* geht es nicht um die Hautfarbe, sondern um eine gesellschaftspolitische Norm und Machtposition. Deshalb wird dieser Begriff in (wissenschaftlichen) Text oft klein und kursiv geschrieben.

TINA*

TINA* steht für trans, inter, nichtbinär und agender und umfasst eine vielfältige Gruppe an Menschen, die nichtbinär cis geschlechtlich sind. TINA* grenzt sich dabei besonders von FLINTA* ab, welches cis Frauen beinhaltet.

Trans

Ein breiter Sammelbegriff für Menschen, die über die traditionellen Geschlechtsgrenzen hinausgehen und solche, dessen Identität nicht zu dem bei der Geburt zugeordneten Geschlecht passt.

Transmaskulin & trans Mann

Menschen, denen bei der Geburt nicht das männliche Geschlecht zugewiesen wurde, die aber männlich oder teilweise männlich sind, können sich als transmaskulin bezeichnen.

Ein Mann, dem bei der Geburt nicht das männliche Geschlecht zugewiesen wurde, ist ein trans Mann.

Quellen des Glossars

- Queer-Lexikon (o.J.): Glossar. Abrufbar unter: <https://queer-lexikon.net/glossar/> [Stand: 24.07.2024].
- Neue deutsche Medienmacher (o.J.): Glossar. Abrufbar unter: <https://glossar.neue-medienmacher.de/glossar/> [Stand: 24.07.2024].
- An alle gedacht?! – GAmM-Broschüre zu Intersektionalität in der Medienpädagogik. Abrufbar unter: www.digitale-chancen.de/materialien/detail/an-alle-gedacht-gamm-broschuere-zu-intersektionalitaet-in-der-medienpaedagogik [Stand: 24.07.2024].
- Pertsch, Sebastian (Hrsg.) (2023): Vielfalt. Das andere Wörterbuch. 100 Wörter – 100 Menschen – 100 Beiträge. Berlin: Dudenverlag.
- Awareness Sankt Pauli (o.J.). Abrufbar unter: <https://awareness-stpauli.de/begriffs-erklarungen/> [Stand 28.08.2024].