

Darüber hinaus würde das Zustimmungserfordernis die Rechtsposition von Müttern schwächen, und so den gesetzgeberischen Wertentscheidungen der Kindschaftsrechtsreform von 1997 zuwiderlaufen. Die Reform sollte die privatautonome Entscheidung der Mutter über die Anerkennung der Vaterschaft fördern, indem die bis dahin erforderliche Zustimmung des Amtspflegers in seiner Funktion als gesetzlicher Vertreter des Kindes zur Vaterschaftsanerkennung ersatzlos aufgehoben wurde. Mit der geplanten Neuerung höhlt die Bundesregierung den Autonomiezuwachs der Mutter aus.³⁰

Fazit

Bereits unter der geltenden Rechtslage zeigt sich eine diskriminierende Verwaltungspraxis, wie das Urteil des OVG Bremen exemplarisch belegt: Behörden treffen pauschalisierende, rasifizierende Annahmen, anstatt einzelfallbezogene Prüfungen

durchzuführen. Die geplanten Gesetzesänderungen der 21. Legislaturperiode würden diese Problematik unverhältnismäßig verschärfen. Eine pauschale Zustimmungspflicht der Ausländerbehörde bei einem sogenannten „Aufenthaltsgefälle“ greift in die Schutzbereiche der Art. 6 Abs. 2 S. 1, Art. 6 Abs. 1 GG, Art. 16 Abs. 1 GG und Art. 3 Abs. 1, Abs. 3 GG ein und gefährdet bestehende Familien. Bereits 2013 stellte das BVerfG klar, dass staatliche Eingriffe in familiäre Bindungen engen verfassungsrechtlichen Grenzen unterliegen. Die Frage bleibt, inwieweit der Rechtsstaat seinem eigenen Anspruch auf Gleichbehandlung und Diskriminierungsfreiheit gerecht wird.

30 Vgl. djb-Stellungnahme 25-36 vom 24.11.2025, S. 1 f., <https://www.djb.de/presse/stellungnahmen/detail/st25-36>.

DOI: 10.5771/1866-377X-2026-1-19

Objektive Künstliche Intelligenz als Mythos – Rechtliche Lücken im Schutz vor algorithmischer Diskriminierung



Carlota Memba Aguado LL.M.
Doktorandin an der juristischen Fakultät der Vrije Universiteit Amsterdam

I. Problemaufriss: Algorithmisch erzeugte Formen der Diskriminierung

Auf künstlicher Intelligenz (KI) basierende Chatsysteme wie *ChatGPT* werden inzwischen von einer Vielzahl von Menschen als integraler Bestandteil des alltäglichen Lebens betrachtet. Der digitale Raum und

die hier zum Einsatz kommenden Algorithmen sind neben technischen Herausforderungen jedoch nicht frei von diskriminierenden Verzerrungen.¹ Studien belegen, dass generative KI diskriminierende Stereotype wiedergibt und verstärkt. Beispielsweise zeigt eine Studie aus dem Jahr 2023, dass die Repräsentation verschiedener Berufsgruppen in von KI-generierten Bildern hinsichtlich Geschlecht und ethnischer Zugehörigkeit nicht mit der durchschnittlichen Selbsteinschätzung der Angehörigen der betreffenden Berufsgruppen übereinstimmt.²

Mit dem Ziel einer besseren empirischen Erfassung des Themas begann ich in Vorbereitung dieses Beitrags einen Selbsttest. Ich erbat *ChatGPT* basierend auf den Konversationen, die wir innerhalb der letzten sechs Monate geführt haben, ein Porträt meiner Person zu erstellen und die Entscheidungen, die das System für die Generierung des Bildes trifft, zu erläutern. Die von *ChatGPT* generierte Darstellung zeigt „eine junge, intellektuelle Frau mit lateinamerikanischem oder südeuropäischem Hintergrund“.³ *ChatGPT* hat mich als eine „Weiß(e), mestiza“⁴ dargestellt, ob-

wohl ich mich nie explizit als solche positioniert habe. Auf die Frage, aus welchem Grund ich nicht als Schwarze⁵ Person, sondern als Weiß dargestellt wurde, gab die chatbasierte KI zur Antwort, dass keine explizite Positionierung als Schwarze Frau erfolgte, und dass ich keine Äußerungen getätigt habe, die vermeintlich „typischerweise mit Schwarzer Sichtbarkeit (...) verbunden sind (z.B. Polizei, Fetischisierung, Haarpolitik, Colorism)“⁶.

Die Antworten deuten auf eine einseitige und stark stigmatisierende Rhetorik hin, die auf stereotypischen Vorstellungen basiert. Dieses Phänomen ist das „Ergebnis eines Modells, das versucht, die wahrscheinlichste Darstellung aus (...) Gesprächen zu konstruieren – und dabei systemisch privilegierte Bilder bevorzugt, wenn kein starker Kontrastimpuls gegeben wird“.⁷

- 1 Unter *Verzerrungen* wird für die Zwecke dieses Beitrages eine subjektive Wahrnehmung oder Neigung verstanden.
- 2 Bianchi, Federico et al.: Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale, in: FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, S. 1493–1504; available at <https://doi.org/mkw9>.
- 3 Zitat aus ChatGPT.
- 4 Zitat aus ChatGPT. Der Begriff „Mestiza“ wird im Kontext der Rassifizierung von Menschen aus Lateinamerika benutzt, um in den geographischen Kontext Weiß gelesene Menschen, die europäische und indigene Wurzeln haben, zu beschreiben. Der Begriff stammt aus der Kolonialzeit.
- 5 „Schwarz“ und „Weiß“ werden in dem Kontext großgeschrieben, um deren sozialen Konstruktionscharakter hervorzuheben und sich von biologischen Kategorien zu distanzieren.
- 6 Zitat aus ChatGPT.
- 7 Zitat aus ChatGPT, als es gefragt wurde, warum es mich nicht als *Afrolatina* darstellte, Hervorhebungen wurden übernommen.

II. Die Architektur des digitalen Raumes

Die im oben dargestellten Beispiel genannten Verzerrungen sind zum einen darauf zurückzuführen, dass KI-Systeme auf Daten trainiert werden, die eine gesellschaftliche Realität abbilden, die selbst von Verzerrungen geprägt ist. Des Weiteren sind diese Verzerrungen auf die Menschen zurückzuführen, die die Algorithmen für die KI erschaffen und deren Perspektiven in die „Architektur des digitalen Raums“ Eingang finden. Algorithmen bestehen grundsätzlich aus vordefinierten „Regeln“ und können darauf basierend selbständig neue „Regeln“ erschaffen. Die Wahrnehmung der Produktentwickelnden spielt vor diesem Hintergrund eine zentrale Rolle.⁸ Eine realitätsgetreue Repräsentation diverser demografischer Gruppen in diesem Bereich könnte zu einer Vielfalt an Perspektiven führen und damit zu weniger diskriminierenden KI-Systemen beitragen. Eine Analyse der Daten der größten Technologieunternehmen, wie beispielsweise *Google*, *Meta* oder *Amazon*, offenbart jedoch signifikante Disparitäten hinsichtlich der Anzahl Schwarzer und Weißer, sowie weiblicher und männlicher Mitarbeitender, insbesondere in den Bereichen der Produktentwicklung.⁹

Diese Verzerrungen können fehlerhafte automatisierte Entscheidungen zur Folge haben. Die potenziell katastrophalen Folgen davon zeigen sich exemplarisch in der „Kindergeld-Affäre“, die sich vor einigen Jahren in den Niederlanden abspielte. Infolge einer von der KI hergestellten Korrelation zwischen u.a. der doppelten Staatsangehörigkeit einer Person und einer erhöhten Wahrscheinlichkeit für Sozialbetrug im System wurden etwa 20.000 Familien qua Verwaltungsakt aufgefordert, das über mehrere Jahre erhaltene Kindergeld zurückzuzahlen.¹⁰ Der Großteil der Betroffenen waren Familien mit sogenanntem „nicht-westlichem Migrationshintergrund“.¹¹

Wie die Kindergeld-Affäre zeigt, kann eine Diskriminierung durch automatisierte Entscheidungssysteme entstehen, indem – mitunter ohne belastbare empirische Grundlage – Korrelationen mit Diskriminierungsmerkmalen hergestellt werden. Die Frage der Kausalität bleibt hierbei außer Acht.¹² Die Anwendung derartiger automatisierter Technologien bei beispielsweise der Kreditwürdigkeitsprüfung, in Bewerbungsprozessen oder in der Versicherungswirtschaft, wird häufig damit begründet, dass sie als frei von menschlicher Einflussnahme und rein datenbasiert gilt. Dadurch entsteht der Anschein objektiver Entscheidungsfindung.

Der Eindruck von Objektivität gründet vor allem auf der vermeintlichen Neutralität von Daten, obwohl diese ebenso durch eine vorherrschend Weiße Perspektive interpretiert werden. Die Idee der Objektivität und Neutralität bedingt eine gesellschaftliche Farbenblindheit, in der angenommen wird, dass sich die Lebensrealitäten aller Menschen gleichen. Von Menschen geschaffene Objekte und Systeme, zu denen auch Systeme Künstlicher Intelligenz gehören, sind demnach nicht neutral oder objektiv.¹³

III. Gesetzliche Anti-diskriminierungsmaßnahmen im digitalen Raum

Der digitale Raum ist ein öffentlicher Ort, in dem Diskriminierung, ebenso wie in physischen öffentlichen Räumen, durch Individuen, Institutionen und Systeme reproduziert wird.

Schutz vor Diskriminierung auf der Ebene der Europäischen Union (EU) wird vor allem durch die Gleichbehandlungsrichtlinien geboten.¹⁴ Diese verbieten jegliche Form der direkten oder indirekten Diskriminierung auf Grund von vor Diskriminierung geschützten Merkmalen in spezifischen Kontexten. Für den Arbeitskontext sieht Art. 1 der Rahmenrichtlinie Beschäftigung ein Verbot jeder „Diskriminierung wegen der Religion oder der Weltanschauung, einer Behinderung, des Alters oder der sexuellen Ausrichtung“ vor. Die Merkmale des Geschlechts und der Rasse oder ethnischen Herkunft sind darüber hinaus durch die Richtlinie zur Gleichstellung der Geschlechter auch außerhalb der Arbeitswelt und die Antirassismusrichtlinie auch im Kontext des Zugangs und Versorgung mit Gütern und Dienstleistungen geschützt. Außerdem schützt Art. 3 der Antirassismusrichtlinie Menschen gegen Diskriminierung aufgrund der ethnischen Zugehörigkeit oder aus rassistischen Gründen auch im Bildungsbereich und in der Sozialversicherung und im Gesundheitswesen. Damit ist die Rasse oder ethnische Herkunft das am umfassendsten geschützte Merkmal.

- 8 Ribeiro Saturnino, Rodrigo: The Colour of Technology. How structural racism is building the digital society, in: Roldão, Cristina et al.: *Afropeans Identities, Racism and Resistance*, London 2025, S. 196–197.
- 9 Laut eigenen Angaben von Meta betrug der Anteil von Frauen in „tech-related positions“ 25,8 %, an Schwarzen Mitarbeitenden 2,4 % in 2022. Bei Amazon betrug der Anteil an Frauen in „corporate positions“ 33 % und an Schwarzen Mitarbeitenden 9,3 % in 2024. Bei Google gab es in 2024 für den EMEA (Europe, Middle-East and Africa) Raum nur 4,1 % Schwarze Mitarbeitende und 34 % Frauen.
- 10 College voor de Rechten der Mens, Vooronderzoek naar de vermeende discriminerende effecten van de werkwijzen van de Belastingdienst/Toeslagen (Voruntersuchung zu den mutmaßlichen diskriminierenden Auswirkungen der Arbeitsweise der Steuerbehörde), September 2022.
- 11 In den Niederlanden wird für statistische Zwecke der Migrationshintergrund in „Western“ und „Non-Western“ aufgeteilt, wobei das letztere laut dem Niederländischen Amt für Statistik Personen meint, die aus einem Land in Afrika, Südamerika, Asien (außer Indonesien und Japan) oder aus der Türkei stammen.
- 12 Antidiskriminierungsstelle des Bundes: *Automatisch Benachteiligt. Das Allgemeine Gleichbehandlungsgesetz und der Schutz vor Diskriminierung durch algorithmische Entscheidungssysteme*, Rechtsgutachten, 2023, S. 23.
- 13 Winner, Langdon: Do Artifacts have Politics? in *Daedalus*, Vol. 109, No. 1, *Modern Technology: Problem or Opportunity?* (Winter, 1980), S. 126, online: <https://www.jstor.org/stable/20024652>.
- 14 Richtlinie zur Festlegung eines allgemeinen Rahmens für die Verwirklichung der Gleichbehandlung in Beschäftigung und Beruf (Rahmenrichtlinie Beschäftigung) (2000/78/EG) Amtsblatt Nr. L 303 vom 02/12/2000 S. 0016–0022, Richtlinie des Europäischen Parlaments und des Rates zur Verwirklichung des Grundsatzes der Gleichbehandlung von Männern und Frauen in Arbeits- und Beschäftigungsfragen (Gender-Richtlinie) (2006/54/EG) Amtsblatt Nr. L 204 vom 05/07/2006 S. 0023–0036, Richtlinie des Rates zur Verwirklichung des Grundsatzes der Gleichbehandlung von Männern und Frauen beim Zugang zu und bei der Versorgung mit Gütern und Dienstleistungen (Richtlinie zur Gleichstellung der Geschlechter auch außerhalb der Arbeitswelt) (2004/113/EG) Amtsblatt Nr. L 373 vom 13/12/2004 S. 0037–0043, Richtlinie zur Anwendung des Gleichbehandlungsgrundsatzes ohne Unterschied der Rasse oder der ethnischen Herkunft (Antirassismusrichtlinie) (2000/43/EG) Amtsblatt Nr. L 180 vom 19/07/2000 S. 0022–0026.
- 15 Verordnung (EU) 2024/1689, Amtsblattversion vom 13. Juni 2024.

Im Bereich der KI besteht durch die jüngst in Kraft getretene EU-Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (kurz: KI-Verordnung)¹⁵ gemäß Art. 5(c) zudem ein Verbot der Verwendung von KI-Systemen, die bei einer Bewertung oder Einstufung einer Person oder Gruppe aus sozialen Zusammenhängen schlechter stellt oder benachteiligt. Die KI-Verordnung bietet auch einen gewissen Schutz vor Diskriminierung, sofern ein KI-System als Hochrisikosystem eingestuft wird. Bei einer solchen Einstufung sind die „Anbieter“ der KI gemäß Art. 9 der KI-Verordnung dafür verantwortlich, das Risikomanagementsystem regelmäßig zu überprüfen und erforderlichenfalls zu aktualisieren. Darüber hinaus haben sie sicherzustellen, dass die weiteren Anforderungen an Hochrisikosysteme (Art. 10–15 der KI-Verordnung) erfüllt werden.

Des Weiteren unterliegen digitale Plattformen, auf denen KI-Systeme häufig zum Einsatz kommen, der EU-Verordnung über einen Binnenmarkt für digitale Dienste (kurz Gesetz über digitale Dienste, DSA).¹⁶ Für sehr große Onlineplattformen¹⁷ besteht gemäß Art. 34 DSA die Verpflichtung, eine Risikobewertung ihrer algorithmischen Systeme in Bezug auf die Ausübung der Grundrechte – u.a. das im Art. 21 der Charta der Grundrechte der EU verankerte Recht auf Nichtdiskriminierung – durchzuführen. Zudem sind sie gem. Art. 35 DSA verpflichtet, Maßnahmen zur Risikominderung zu ergreifen. Aktuell wurden in der gesamten EU 22 sehr große Online-Plattformen i.S.d. DSA identifiziert, darunter Netzwerkplattformen wie *Facebook* und *X* (ehemals *Twitter*), Suchmaschinen wie *Google*, aber auch Plattformen für Erwachseneninhalte, wie *Pornhub*.

In Deutschland bietet grundsätzlich das Allgemeine Gleichbehandlungsgesetz (AGG) Schutz vor Diskriminierung. Für den digitalen Raum ist dieser Schutz jedoch nur eingeschränkt wirksam. Es fehlen konkrete Diskriminierungstatbestände für digital vermittelte Benachteiligungen, und auch die Rechtsdurchsetzung ist unzureichend, wie ein Gutachten der Antidiskriminierungsstelle des Bundes aus dem Jahr 2023 zeigt.¹⁸

IV. Fazit

Die Rechtslage *de lege lata* gewährleistet gegenwärtig keinen effektiven Schutz gegen algorithmenbasierte Diskriminierung und sieht auch keine Mechanismen vor, um dieser Form der Diskriminierung präventiv zu begegnen. Der unionsrechtliche Schutz vor Diskriminierung ist unzureichend. Der DSA verpflichtet lediglich sehr große Plattformen zu Risikobewertungen. Der

Anwendungsbereich der allgemeinen Gleichbehandlungsrichtlinien umfasst zwar auch den Zugang zu Dienstleistungen, sodass von dem Anwendungsbereich dieser auch kleinere Onlineplattformen erfasst werden können. Die Richtlinien beziehen sich jedoch lediglich auf die geschlechtsbezogene und rassistische Diskriminierung. Der Ansatz der manuellen Überprüfung und eventuellen Modifikation der Algorithmen künstlicher Intelligenz nach Art. 9 des KI-Gesetzes bietet keinen ausreichenden Schutz vor Diskriminierung; vor allem ermöglicht er keine präventiven Maßnahmen. Einer effektiven Rechtsdurchsetzung stehen häufig bereits Informationsasymmetrien und die Intransparenz algorithmischer Systeme im Wege.

Der unzureichende Schutz vor Diskriminierung ist jedoch kein rein digitales Problem. Wie bereits dargelegt, beruht die KI-generierte Diskriminierung im Wesentlichen auch darauf, dass bestehende gesellschaftliche Ungleichheiten in den Datengrundlagen und -strukturen abgebildet und zum Teil verstärkt werden. Eine wirksame Bekämpfung digitaler Diskriminierung setzt daher voraus, dass Diskriminierung auch außerhalb des digitalen Raums konsequent adressiert wird.

Erforderlich sind gesetzliche Maßnahmen, die Diskriminierung nicht nur auf der individuellen zivilrechtlichen Ebene untersagen, sondern darüber hinaus auch systemische und institutionelle Formen erfassen. Ein Ansatz könnte darin bestehen, das geltende Recht so zu erweitern, dass auch Konstellationen wie diskriminierende Praktiken staatlicher Exekutivorgane einbezogen werden.

Darüber hinaus kann eine verstärkte Inklusion sowie eine größere Diversität des Personals in der Produktentwicklung der KI-Systeme dazu beitragen, die Reproduktion diskriminierender Stereotype durch generative KI-Systeme zu verringern. Der gesetzliche Antidiskriminierungsschutz sollte verstärkt auf präventive Maßnahmen setzen, etwa durch verpflichtende Förderprogramme für benachteiligte Gruppen.

16 Verordnung (EU) 2022/2065, Amtsblatt Nr. L 277 vom 19/10/2022, S. 0001–0102.

17 Gem. Art. 33 des Gesetzes über digitale Dienste werden unter sehr großen Onlineplattformen diejenigen Plattformen verstanden, die eine durchschnittliche monatliche Zahl von mind. 45 Millionen aktiven Nutzer*innen in der Union haben.

18 Antidiskriminierungsstelle des Bundes: *Automatisch Benachteiligt. Das Allgemeine Gleichbehandlungsgesetz und der Schutz vor Diskriminierung durch algorithmische Entscheidungssysteme*, Rechtsgutachten, 2023.