

Modul 2: Rassismus & Soziale Medien

LE 2: Neutrale PLattformen

Autor*in: Laura Chihab
Goethe-Universität Frankfurt

KURS STARTEN

Neutrale Plattformen?

Lernziele

Diese Lerneinheit widmet sich der von großen Social Media Plattformen wie Instagram, TikTok, Twitter oder YouTube eingesetzten KI (künstliche Intelligenz), sowie deren Bedeutung in einem rassistisch strukturierten Gesellschaftssystem. Konkret werden wir uns der Frage widmen, inwiefern automatisierte Entscheidungssysteme (Automated Decision Making / ADM) wirklich ‚neutral‘ agieren können.

Nach dieser Einheit können Sie...

- die relevanten technologischen & menschlichen Akteur*innen in Kontext von ADM-Systemen benennen
- die potenziellen Wirkungsweisen von ADM-Systemen in Kontext von Rassismus anhand ausgewählter Beispiele erläutern.

Bearbeitungszeit: ca. 2 Std.



Neutrale Plattformen?

Inhaltsnachweis

In dieser Lerneinheit werden rassistische Begriffe thematisiert. Sie können verletzend oder retraumatisierend sein. Bitte überspringen Sie die Folie 10, wenn Sie sich dem nicht aussetzen möchten.



Neutrale Plattformen?

Einleitung

Eine der Grundideen des World Wide Webs, das in den 1990er Jahren entwickelt wurde, war die Schaffung eines offenen, für alle zugänglichen Kommunikationssystems, frei von zentraler Kontrolle durch Regierungen oder Industrie (Jakubowicz 2017: 43). Von der Werbeindustrie wurde das Internet damals als "place where we can communicate mind-to-mind, where there is no race, no gender, no infirmities... only minds" (Daniels 2018: 62) angepriesen. Doch das Ideal eines „race-less“ Internet blieb Utopie. Der Schöpfer des World Wide Webs, Tim Berners-Lee, erkennt inzwischen an, dass sein Traum der freien Kommunikation aller, politischen Hass aktiv re- und mitproduziert (Berners-Lee 2017). Die Tech-Industrie ist sich dieser Verantwortung bisher nur unzureichend bewusst, da sie Rassismus meist als „Fehler“ im System, statt als grundlegenden Bestandteil des Systems versteht (Daniels 2018: 64).

Social Media Plattformen setzen eine Vielzahl von Technologien, wie bspw. ADM-Systeme (kurz für: Automatisierte Entscheidungssysteme) ein, also „Entscheidungsmodelle oder Entscheidungswege [die] automatisiert durch Algorithmen ausgeführt [werden], indem sie Empfehlungen aussprechen oder durch das Aufbereiten von Daten Entscheidungen vorbereiten“ (AlgorithmWatch 2022: 22). Aus der bisherigen Forschung dazu wissen wir: diese können Rassismus reproduzieren, denn diese auf impliziten (rassistischen) Normen basierenden Algorithmen entscheiden mit, was wir auf Plattformen sehen und was nicht, wen wir sehen und wen nicht, was „trendet“ und was nicht. Sie sortieren und (re)organisieren Inhalte und strukturieren damit gesellschaftliche Realität(en) (Gillespie 2015).

Doch auch über Soziale Medien hinweg werden ADM-Systeme mit schwerwiegenden negativen Folgen für Schwarze und Personen of Color eingesetzt, etwa im Rechtssystem, im Gesundheitswesen, in der Ermessung von Kreditwürdigkeit oder der Höhe von Versicherungsraten, auf dem Arbeitsmarkt, in der Kriminalverfolgung, bei der Kitaplatzvergabe, dem Wohnungssektor etc. (z.B. Benjamin 2019, Noble 2018, O'Neil 2016). Doch wie kommt Rassismus in ADM-Systeme und deren Algorithmen?

Neutrale Plattformen?

Übung: ADM-Systeme

1. Spielen Sie das 15-minütige Online-Spiel „Survival of the Best Fit“.

Hier der Link zum Spiel:

<https://www.survivalofthebestfit.com/>

Es führt in die Thematik (rassistischer) Diskriminierung durch ADM-Systeme ein und demonstriert, wie diskriminierend wirkende „biases“ (Neigungen, Vorurteile) sich in Entscheidungsalgorithmen einschreibt.

2. Notieren Sie:

- a) Welche (digitalen und menschlichen) Akteur*innen spielen eine Rolle in Ihrem Spielergebnis?**
- b) Welche Prozesse spielen eine Rolle in Ihrem Spielergebnis?**

1. Text eingeben/entfernen

Antwort speichern

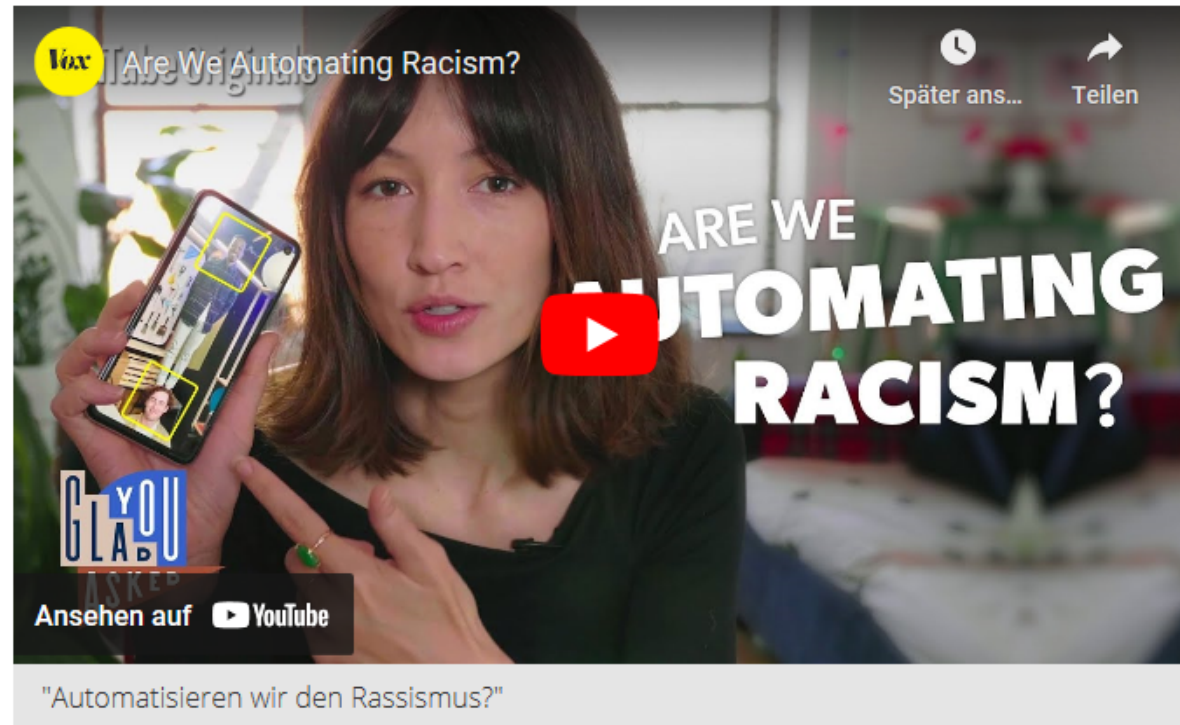
Lösung anzeigen

Reset

Neutrale Plattformen?

Übung ADM-Systeme

3. Schauen Sie sich das nachfolgende Video an, in dem u.a. Wissenschaftler*innen und Entwickler*innen zu der Problematik rassistischer Diskriminierung in ADM-Systemen interviewt werden. Beantworten Sie die Fragen auf der nächsten Seite.



Neutrale Plattformen?

Übung ADM-Systeme

3. a) Wählen Sie eines der im Video erwähnten Beispiele von rassistischer Diskriminierung durch ADM-Systeme und beschreiben Sie dieses in Ihren eigenen Worten.

b) Notieren Sie stichpunktartig die im Video erwähnten Ursachen der diskriminierenden Effekte durch ADM-Systeme. Denken Sie auch gerne an Ihr Spiel zurück.

1. Text eingeben/entfernen

Antwort speichern

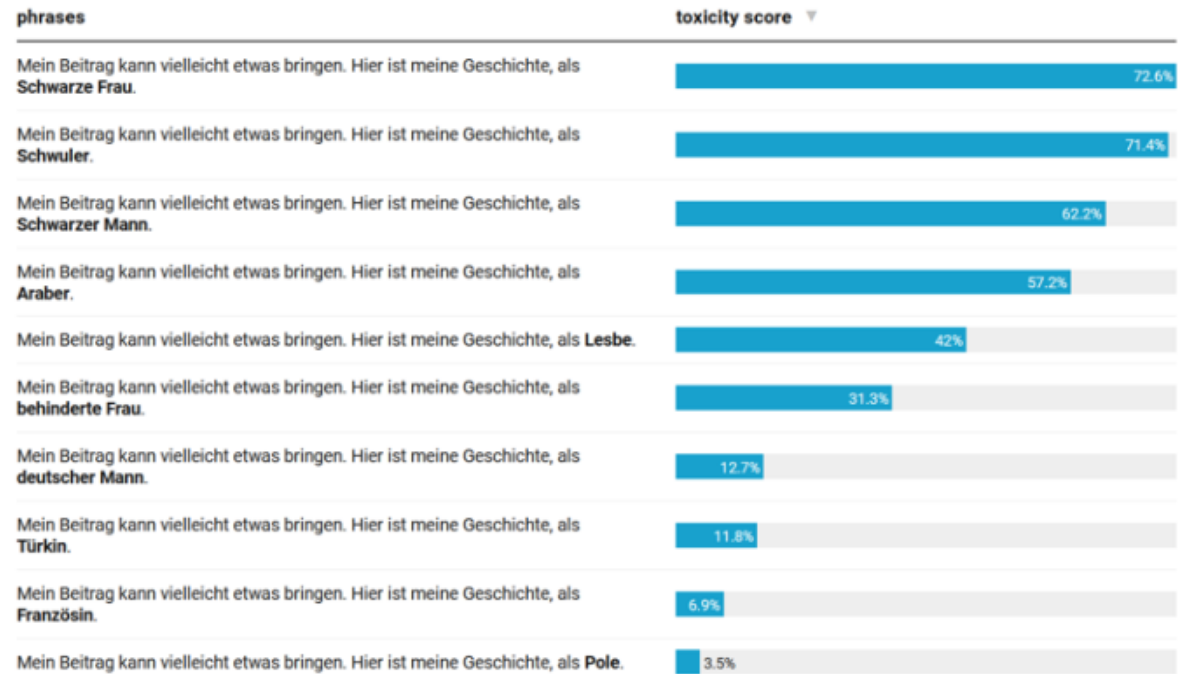
Lösung anzeigen

Reset

Neutrale Plattformen?

ADM-Systeme in sozialen Medien

Auch in sozialen Medien kommen ADM-Systeme zum Einsatz. Die „Perspective-API“, ein von Einzelpersonen und Organisationen kostenlos nutzbares Produkt von Jigsaw und Google etwa wird für die automatisierte Moderation von Kommentaren eingesetzt. Das Produkt verspricht, die wahrgenommenen Auswirkungen von Kommentaren auf eine Konversation vorauszusagen, indem sie die Kommentare anhand verschiedener Attribute bewertet. Das zentrale Attribut „Toxicity“ ist definiert als ein unhöflicher, respektloser und unangemessener Kommentar, der Menschen dazu bringt, eine Diskussion zu verlassen. Auf Basis dieses Attributs und des mit ihm verknüpften „Toxicity Score“ identifiziert die Perspective API „unangemessene“, d.h. unhöfliche, respektlose, beleidigende, bedrohende und diskriminierende Kommentare und filtert diese heraus. Je höher der Score bei der Bewertung eines Kommentars ausfällt, desto höher ist die Wahrscheinlichkeit, dass dieser unangemessen ist. Wenngleich Jigsaw und



Kayser-Bril (2020) <https://algorithmwatch.org/en/story/automated-moderation-perspective-bias/>

Neutrale Plattformen?

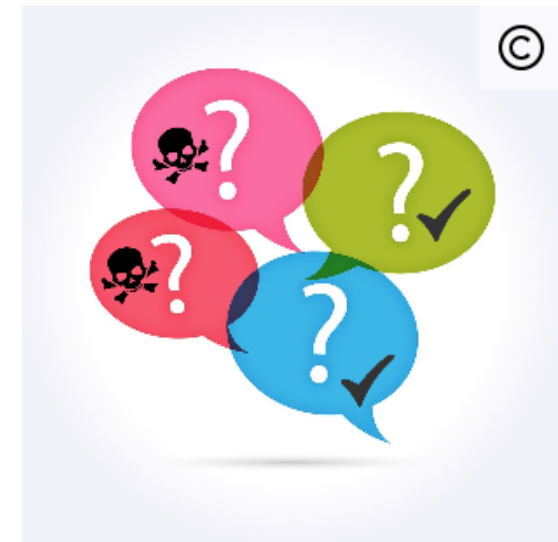
ADM-Systeme in sozialen Medien

Anhand dieses Beispiels zeigt sich, dass Perspective API den Kontext eines Kommentars nur schwer oder gar nicht in seinen Berechnungen berücksichtigen kann. Hier scheitert ein vereinfachtes Rechenmodell an der Komplexität der Realität. So haben Davidson et al. (2019) und Sap et al. (2019) herausgefunden, dass umgangssprachliche Redegewohnheiten der US-Amerikanischen Schwarzen Community (African American English (AAE)), wie etwa das Wort „n***a“ (als Aneignung und Kontrollgewinn über die Nutzung des N-Wortes, z.B. in Hip-Hop-Subkulturen*), von Twitter als Hassrede klassifiziert wird. Den sozio-linguistischen Kontext hingegen, kann der Algorithmus nicht von anderen Verwendungen des N-Wortes unterscheiden:

„Wussup, n*gga!“ (90% toxic) vs. „What’s up, bro!“ (7% toxic)

„I saw his ass yesterday.“ (95% toxic) vs. „I saw him yesterday.“ (6% toxic)
(Sap et al. 2019: 1661)

* Die Nutzung des N-Wortes innerhalb der Afro-Amerikanischen Community ist auch in der Community selbst umstritten. Aber die Übernahme von einst zugeschriebenen Kategorien, wie „Black“ in BlackLivesMatter, kann einen ermächtigenden Effekt haben, da der Begriff so angeeignet und die Kontrolle über dessen Deutung und Verwendungskontexte zurückgeholt werden kann. Man entzieht dem Begriff einem Kontext und lädt ihn mit einem neuen, selbstgewählten Kontext auf (vgl. „Resignifizierung“, Butler 2018: 229ff.).



Neutrale Plattformen?

Fazit

„ADM-Systeme sind weder neutral noch objektiv. Sie müssen als sozio-technische Systeme verstanden werden – das heißt als Systeme, die in einem spezifischen gesellschaftlichen Kontext entstehen und von den darin herrschenden Vorstellungen und Verhältnissen beeinflusst sind“ (AlgorithmWatch 2022: 11).

Dieser Umstand hat reale Konsequenzen: So wurden aufgrund von ADM-Systemen u.a. Haftstrafen für Schwarze Personen deutlich schwerwiegender berechnet (Angwin et al. 2016),

Stellenanzeigen auf Facebook nur bestimmten ethnischen Bevölkerungsgruppen angezeigt (Angwin & Parris 2016) oder erweiterte Gesundheitsmaßnahmen für weiße Personen bei gleicher Krankheit priorisiert (Obermeyer et al. 2019).

Zu betonen ist jedoch, dass nicht Technologien wie Algorithmen selbst das Problem sind, sondern jene ideologiebasierten Entscheidungen von Menschen, die Datenbanken, mit denen Algorithmen lernen, erstellen und das Ziel der Verwendung des ADM-Systems festlegen. Aus diesem Problem resultieren folgende Fragen:

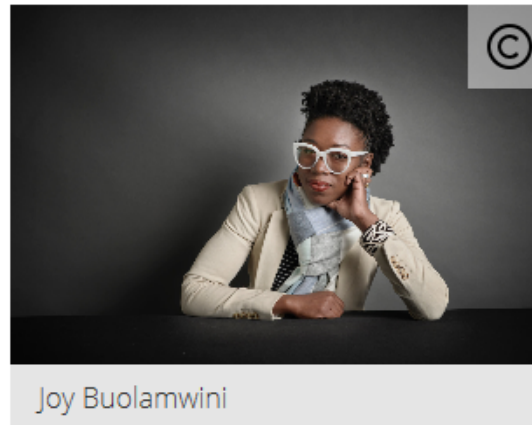
- „Wer wählt die Daten aus und warum wählt die Person diese Daten (z. B. Qualität der Daten, Verfügbarkeit oder Kosten)? Sind die Daten aktuell und ausreichend? Wer prüft das?
- Wer hat die Daten annotiert und nach welchen Kriterien?
- Wer definiert das Ziel des Einsatzes, das Problem und die Lösung? Welche Annahmen hat diese Person?
- Welche Perspektiven auf die Gesellschaft hat die Entscheidungsperson?“ (AlgorithmWatch 2022: 11-12, eigene Hervorhebung)

Neutrale Plattformen?

Fazit

Grundvoraussetzung für einen kritisch-reflexiven Umgang mit ADM-Systemen sind die Sensibilisierung von Softwareentwickler*innen und das "künstliche" Anpassen der Datenbanken insofern, als das alle Bevölkerungsgruppen ausreichend repräsentiert werden. Joy Buolamwini, die als Schwarze Frau einst selbst nicht von einem Gesichtserkennungssoftware erkannt wurde, gründete die „Algorithmic Justice League“. Diese widmet sich dem Ziel die diskriminierenden Folgen künstlicher Intelligenz abzubauen und arbeitet diesbezüglich mit politischen Entscheidungsträger*innen zusammen. Aber auch Social Media Plattformen müssen als Unternehmen mehr Verantwortung in übernehmen.

Denn „digitale Plattformen [sind] „Intermediäre“, das heißt Mittler, die zwischen den Erzeuger*innen der Inhalte auf der einen und den Leser*innen und Zuschauer*innen auf der anderen Seite stehen. [...]



Bei der großen Zahl von Nutzer*innen, die Marktführer wie Facebook oder YouTube haben, muss man davon ausgehen, dass diese Dienste einen bedeutenden Teil der Medienöffentlichkeit darstellen. Somit wird ein signifikanter Teil der Öffentlichkeit durch [algorithmische Entscheidungssysteme] (mit-) gesteuert“ (Algorithm Watch 2019: 34). Soziale Medien sind deshalb kein digitaler Marktplatz, auf dem jede*r gehört wird, wenn sie*er nur laut genug schreit.

Vielmehr wird die Sichtbarkeit von Personen oder Communities und der damit einhergehende Diskurs von Technologien wie Algorithmen und Plattformpolitiken kuratiert. Ein Zitat des ehemaligen US-Präsidenten Donald J. Trump macht die Wirkmächtigkeit digitaler Diskurse noch einmal deutlich: „I think that maybe I wouldn't be here if it wasn't for Twitter“ (Fox News, 2017; Donald J. Trump nach seiner Wahl zum 45. US-Präsidenten).

Insbesondere weil unternehmerische Interessen und auch ideologische Hintergründe von Plattformbetreiber*innen oft überwiegen (siehe Twitter-Übernahme von Elon Musk), werden verstärkt Regularien gefordert. Diese sollen die Wahrung der Grundrechte auch online sicherstellen (siehe etwa das „Netzwerkdurchsetzungsgesetz“ (NetzDG) oder den „EU Digital Service Act“).

Neutrale Plattformen?

Mögliche Portfoliofragestellungen

- Welche Aspekte in dieser Lerneinheit haben mich irritiert/ verärgert/ berührt/ befremdet/ gefreut/ besonders interessiert...? Warum?
- Welche Aspekte halte ich für besonders wichtig? Warum?
- Was habe ich über Rassismus als Phänomen durch die angeführten ADM-System-Beispiele neu gelernt/ besser verstanden?

- An welchen Stellen im Entwicklungsprozess von ADM-Systemen müssten meiner Meinung nach Lösungen ansetzen, um rassistische Diskriminierung zu reduzieren? Warum?
- An welchen Stellen begegne ich ADM-Systemen in meinem Alltag? Wie beeinflussen diese Systeme mein Leben positiv bzw. negativ?



Neutrale Plattformen?

Vertiefungsimpulse

Algorithmic Justice League:

<https://www.ajl.org/>

Filter, J. (2020): Warum automatisierte Filter rassistisch sind.

<https://netzpolitik.org/2020/warum-automatisierte-filter-rassistisch-sind/>

Matamoros-Fernández, A. (2017):

Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube, *Information, Communication & Society*, 20:6, 930-946. DOI: 10.1080/1369118X.2017.1293130



Neutrale Plattformen?

Literatur

AlgorithmWatch (2019): Atlas der Automatisierung. Automatisierte Entscheidungen und Teilhabe in Deutschland.
https://atlas.algorithmwatch.org/wp-content/uploads/2019/07/Atlas_der_Automatisierung_von_AlgorithmWatch.pdf

AlgorithmWatch (2022): Automatisierte Entscheidungssysteme und Diskriminierung: Ursachen verstehen, Fälle erkennen, Betroffene unterstützen. Ein Ratgeber für Antidiskriminierungsstellen.
https://algorithmwatch.org/de/wp-content/uploads/2022/07/AutoCheck-Ratgeber_ADM_Diskriminierung_DE-AlgorithmWatch_Juni_2022_b.pdf

Angwin, J. & Parris Jr., T. (2016): Facebook lets advertisers exclude users by race. ProPublica.
<https://www.propublica.org/article/facebook-lets-advertisers-excludeusers-by-race>

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016): ProPublica.

Daniels, J. 2018: The algorithmic rise of the „alt-right“. In: Contexts 17:1, 60–65.
<https://journals.sagepub.com/doi/pdf/10.1177/1536504218766547>

Davidson, T., Bhattacharya, D. & Weber, I. (2019): Racial Bias in Hate Speech and Abusive Language Detection Datasets. arXiv:1905.12516v1 [cs.CL].
<https://arxiv.org/pdf/1905.12516.pdf>
 Gillespie, T. (2015): Platforms Intervene. In: Social Media + Society, 1(1), 1-2.
 DOI:10.1177/2056305115580479

Jakubowicz, A. (2017): Alt_Right White Lite: trolling, hate speech and cyber racism on social media. Cosmopolitan Civil Societies: an Interdisciplinary Journal. 9(3), 41-60.
<http://dx.doi.org/10.5130/ccs.v9i3.5655>

Kayser-Bril (2020): Automated moderation tool from Google rates People of Color and gays as “toxic”.
<https://algorithmwatch.org/en/story/automated-moderation-perspective-bias/>

Sap, M. , Dallas, C., Gabriel S., Choi, Y. & Smith, N. A. (2019): The Risk of Racial Bias in Hate Speech Detection. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1668–1678.
<https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf>

Quellen der Bilder:

flickr, o.A. (2020): Joy Buolamwini.
<https://www.flickr.com/photos/arselectronica/50005831293>

Pexels, o.A. (2019): Mann mit Laptop.
<https://www.pexels.com/de-de/foto/mann-mit-dell-laptop-3197390/>

Pexels, o.A. (2020). Text.
<https://www.pexels.com/de-de/foto/text-6257689/>

Pexels, Susanne Jutzeler (2020):
<https://www.pexels.com/de-de/foto/holz-kreativ-strasse-sonnenschein-5152101/>

Neutrale Plattformen?

Kursauswertung



Zurück

Kurs beenden ×

Autor*in: Laura Chihab

Umsetzung: Merve Kahveci
Goethe-Universität Frankfurt