Digitaler Hass

„Digitale Hassreden“ und Verschwörungsideologien in Zeiten der COVID-19 Pandemie

Digitaler Hass untersucht Hassreden und Verschwörungsnarrative im Netz, die insbesondere im Kontext der COVID-19 Pandemie eine rasche Zunahme erfahren haben. Ein besonderer Fokus wird dabei auf die Verbreitung von rassistischen und antisemitischen Beiträgen, Kommentaren und Aufrufen gelegt, die mit Berliner Kooperationspartner*innen systematisch und praxisnah untersucht wurden.

Dezember 2023
Berlin

Herausgegeben von

Alice Salomon Hochschule Berlin
Alice-Salomon-Platz 5, 12627 Berlin
Vertreten durch Prof. Dr. María do Mar Castro Varela

Haus der Kulturen der Welt
John-Foster-Dulles-Allee 10, 10557 Berlin
Vertreten durch Eva Stein und Naïma Walter

Redaktion und Konzeption

María do Mar Castro Varela
Verónica Orsi

Texte

María do Mar Castro Varela
Helena Mihaljević
Puneh Abdi
Christina Hübers
Monika Hübscher
Bahar Oghalai
Verónica Orsi
Milena Pustet
Elisabeth Steffen

Illustrationen

Hamed Eshrat

Gestaltung

Verónica Orsi

Ein Projekt von

Alice Salomon Hochschule Berlin
University of Applied Sciences

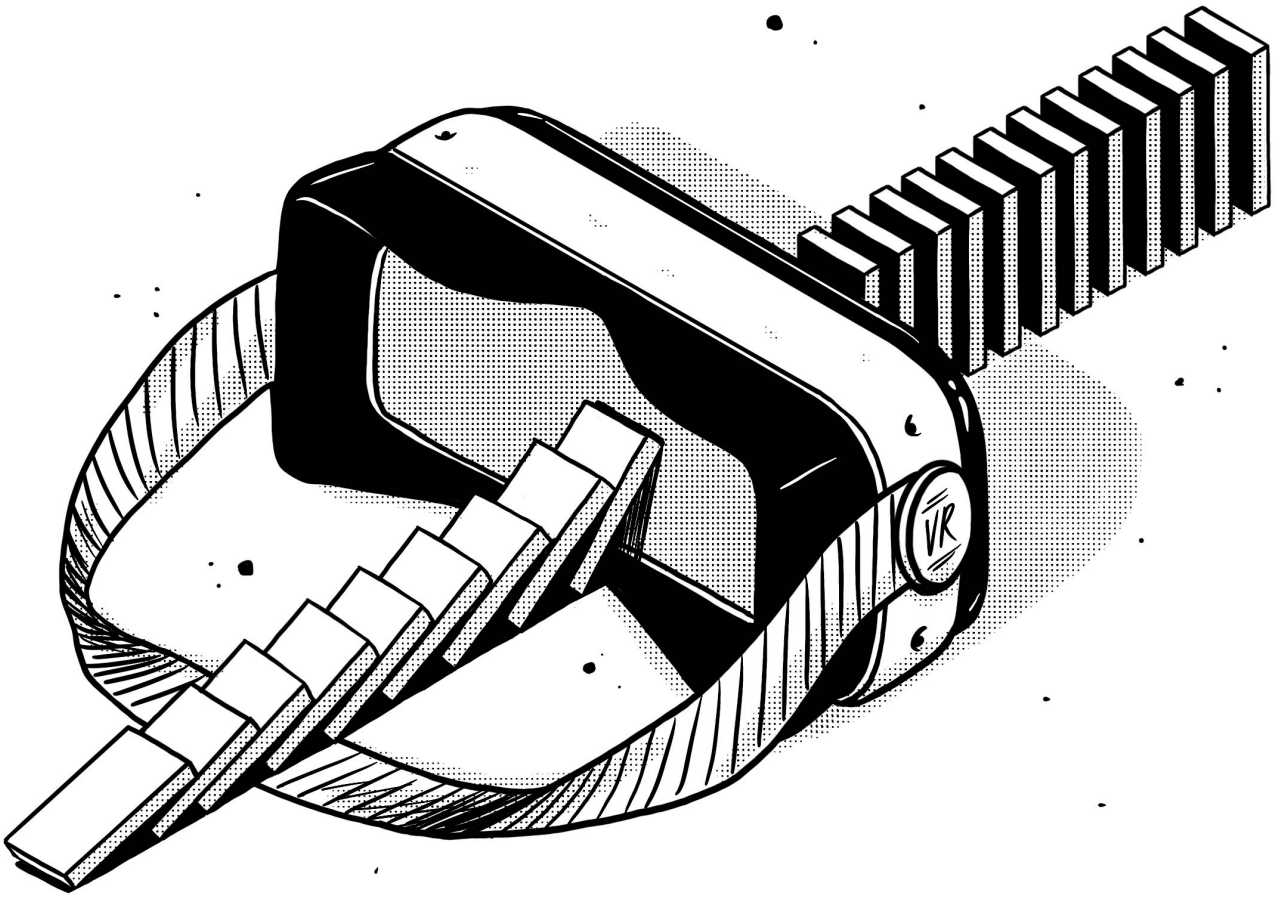
Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences

Haus der Kulturen der Welt

Gefördert durch

BERLIN IFAF

Die Beauftragte der Bundesregierung für Kultur und Medien


VR

Inhalt

Vorwort

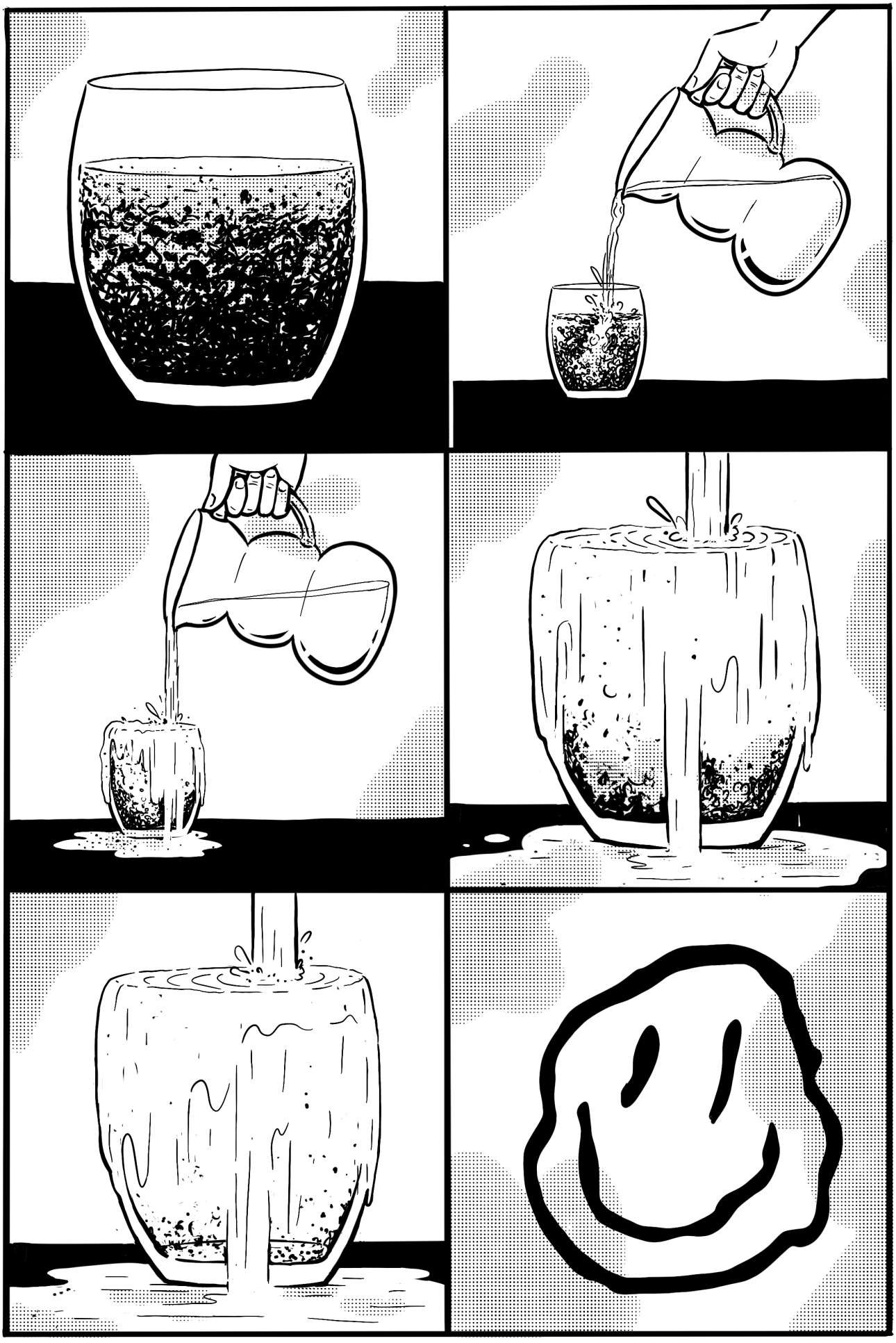
María do Mar Castro Varela

Das Projekt „Digitaler Hass“ hat zwischen 2021 und 2023 Hassreden und Verschwörungsnarrative im Netz untersucht, die insbesondere im Kontext der COVID-19 Pandemie eine rasche Zunahme erfahren haben. Finanziert vom Institut für angewandte Forschung (IFaF) in Berlin wurde ein besonderer Fokus auf die Verbreitung von rassistischen und antisemitischen Beiträgen, Kommentaren und Aufrufen gelegt, die mit Berliner Kooperationspartner*innen diskutiert und untersucht wurden. Geleitet wurde das Projekt von María do Mar Castro Varela (Alice Salomon Hochschule) und Helena Mihlajević (Hochschule für Technik und Wirtschaft). Unter Verwendung von rechnergestützten und algorithmischen Verfahren des Text Mining und des maschinellen Lernens wurden Hassreden und Verschwörungsnarrative in deutschsprachigen Internet-Diskussionen identifiziert, um dann über eine diskursanalytische Feinanalyse den Zusammenhang von Verschwörungsnarrativen mit Rassismus und Antisemitismus genauer herzustellen. Die analysierten Datensätze sollten wiederum für die Verbesserung von Algorithmen zur automatischen Identifikation von digitalem Rassismus und Antisemitismus aufbereitet werden. Die Ergebnisse der Analysen wurden mit zivilgesellschaftlichen Akteur*innen aus der Region, die etwa im Bereich der Antidiskriminierung und der politischen Bildung tätig sind, diskutiert. Diese unterstützen das Projekt bei der Entwicklung bzw. Verbesserung von Maßnahmen zur Entdeckung, Bewertung und insbesondere Bekämpfung digitaler Hassreden und Verschwörungsnarrative. Einige Vereine und Organisationen arbeiten schon seit vielen Jahren an der Frage, wie auf Verschwörungstheorien und digitalen Hass geantwortet werden kann. Die COVID-19 Pandemie hat die Verbreitung von Hassreden und Verschwörungstheorien allerdings enorm beschleunigt. Die Kanäle und Plattformen der sozialen Medien erwiesen sich als nicht resistent gegen eine schnelle Dissemination von Hass und insbesondere gegen Verschwörungen, die staatsphobische Inhalte transportierten.

Der transdisziplinäre Ansatz der Forschung erwies sich als produktiv und herausfordernd zugleich. Die rechnergestützten Techniken ermöglichten es, die schiere Masse von Hassreden und Verschwörungstheorien zu erheben und so sichtbar zu machen. So konnte in der Feinanalyse ein genauer Blick auf die Formen und Inhalte der Diskurse geworfen werden. Wie Algorithmen jedoch rassistischem und antisemitischem Hass begegnen können, ist, wie auch viele andere internationale Untersuchungen gezeigt haben, nicht ganz so einfach. Verschwörungstheorien, die über Verschwörungsfragmente auf Twitter (heute X) verbreitet werden, operieren häufig mit sich ständig wandelnden Codes und Bildern, die nicht so einfach detektierbar sind, während die zivilgesellschaftlichen Debatten um Rassismus und Antisemitismus großen Einfluss darauf haben, wie bestimmte Tweets, Bilder oder Nachrichten auf Telegram bewertet werden. Die Bekämpfung von Hassreden im Internet ist in der Tat eine komplexe Herausforderung. Algorithmen können eine Rolle bei der Identifizierung und Eindämmung von Hassreden spielen, sind jedoch nicht die alleinige Lösung. Algorithmen können Texte analysieren und automatisch nach Hassreden oder problematischen Inhalten suchen. Dies erfordert maschinelles Lernen und NLP (Natural Language Processing) Techniken, um Hassrede von anderen Arten von Text zu unterscheiden. Doch ist es nicht immer einfach und eindeutig zu entscheiden, ob es sich um eine Hassrede oder lediglich um eine „andere Meinung” handelt, die geschützt werden muss. Algorithmen können lange Listen von Schimpfwörtern, rassistischen Ausdrücken und anderen problematischen Begriffen verwenden, um Texte zu filtern oder zu markieren, die solche Wörter enthalten. Sie können zudem das Verhalten von Benutze*innen in sozialen Netzwerken überwachen und unangemessenes Verhalten, wie das wiederholte Posten von Hassreden, erkennen. Insoweit können Algorithmen gemeldete Hassreden schneller identifizieren und so eine schnellere Moderation ermöglichen. Allerdings arbeiten sie nie fehlerfrei, so dass eine menschliche Überprüfung und Moderation notwendig bleiben, um sicherzustellen, dass keine Meinungsfreiheit oder legitimen Diskussionen beeinträchtigt werden. Die Diskursanalyse erwies sich als sinnvoll, um freizulegen wie ideologische Fragmente in den Mitteilungen zusammengeschweißt wurden, so dass frühe Verschwörungstheorien - wie etwa das antisemitische Machwerk „Die Protokolle der Weisen von Zion“ - mit Verschwörungstheorien im Zusammenhang mit der COVID-19 Pandemie wie aus einem Guss erscheinen konnten. Deutlich wurde auch die Kontinuität von antisemitischen und rassistischen Bildern, die geradezu ungebrochen seit hunderten Jahren fortleben: etwa der „gierige Jude“ oder die „gelbe Gefahr“.

In der vorliegenden Broschüre sind kurze Texte versammelt, die zum Nachdenken anregen sollen. Sie liefern kurze Impulse, die dazu animieren, gemeinsam nach Wegen zu suchen, soziale Medien sinnvoll und demokratietauglich zu nutzen. Eine Welt ohne soziale Medien ist nicht mehr möglich, die Frage ist nun, wie soziale Medien für eine bessere Zukunft genutzt werden können und wie Hass im Netz begegnet werden kann.

Berlin, September 2023



„Common Sense"

Puneh Abdi

Verschwörungstheorien neue und alte

In der Wissenschaft sind wir immer wieder mit Fragen der Begriffsklärungen und Definitionen konfrontiert. Gerade in den Hochphasen der COVID-19 Pandemie haben Medien und Politik mit den Begriffen *Verschwörungstheorie*, *Verschwörungsideologie*, *Fake News*, *Desinformation*, etc. um sich geworfen. Schnell führte dies zu Verwirrung, Überforderung oder auch der falschen Benutzung dieser Begriffe. Das Forschungsprojekt *Digitaler Hass* hat nicht nur die Inhalte solcher Verschwörungserzählungen während der COVID-19 Pandemie analysiert, sondern sich auch gründlich mit den jeweiligen Begrifflichkeiten und ihren Bedeutungen beschäftigt. Es soll hier jedoch nicht bloß eine weitere Broschüre mit eigenen Definitionen entstehen. Viele Bildungsträger, Nichtregierungsorganisationen, Vereine und Stiftungen klären über Verschwörungstheorien, Desinformation, Fake News etc. auf. So auch zahlreiche unserer Kooperationspartner*innen. Wir fragen uns deshalb zunächst: Was ist eigentlich Common Sense bezüglich dieser Begriffe? Gibt es überhaupt Einigungen darüber, wie diese Begriffe definiert werden? Welche aktuell laufenden Debatten gibt es, wo wird gestritten?

Verschwörungstheorie, Desinformation, Fake News?

In den Medien kursieren immer wieder Begriffe wie *Desinformation*, *Misinformation* oder auch *Fake News*. Diese werden häufig synonym verwendet, dabei unterscheiden sie sich in einigen wesentlichen Punkten. Wenn wir von einer unbestätigten Information sprechen, zu der es noch keine Angaben über ihren Wahrheitsgehalt gibt, sprechen wir zunächst einmal von einem *Gerücht*. Ist eine Information falsch und wird absichtlich verbreitet, sprechen wir von *Desinformation*. Die unbeabsichtigte Verbreitung einer falschen Information ist wiederum eine *Misinformation*. Der Begriff *Fake News* wird verwendet, um erfundene Nachrichten zu bezeichnen - wir haben aber in den letzten Jahren auch gesehen, wie er immer häufiger verwendet wird, um Nachrichtenmedien zu delegitimieren und als unseriös zu bezeichnen.

Alles nur Verschwörung? Die Begriffe im Überblick

Am gängigsten ist der Begriff der *Verschwörungstheorie*. Dieser wird jedoch immer häufiger kritisiert, da die Endung des Wortes -„Theorie"- irreführend sein kann. Sie impliziert, dass es eine theoretische Grundlage für die Aussagen gibt und dass das in der Theorie Gesagte wissenschaftlichen Standards entspricht. Der Begriff ist trotzdem weit verbreitet und wird auch von Institutionen, die ihn eigentlich kritisieren, weiterhin benutzt, um das Thema zu behandeln. Wenn wir über die Inhalte von Verschwörungstheorien sprechen, benutzen wir den Begriff *Verschwörungserzählung*. Unter dem Begriff können wir verschiedene Erzählungen von Verschwörungsideologen bündeln, in einen zeitlichen Kontext setzen und auch voneinander unterscheiden. Eine Verschwörungsideologie sind diese Erzählungen, wenn sie sich in einem geschlossenen Weltbild manifestieren. Das heißt, selbst wenn es klare Fakten und Belege dafür gibt, dass eine Verschwörungserzählung nicht der Wahrheit entspricht, wird sie aufrechterhalten. Wer ein verschwörungsideologisches Weltbild hat, lässt jeglichen Widerspruch abprallen und glaubt trotzdem an seine Erzählung.

Merkmale von Verschwörungstheorien

Verschwörungstheorien versuchen, eine alternative Realität und Interpretation von politischen, gesellschaftlichen oder medialen Ereignissen herzustellen. Dabei sehen wir bestimmte Merkmale, die in verschiedenen Verschwörungserzählungen immer wieder auftauchen. Verschwörungstheorien bieten einfache Antworten auf unübersichtliche, komplexe soziale und politische Ereignisse. In ihnen wird der Glaube formuliert, dass es im Geheimen agierende Mächte gibt, die einen bösen Plan verfolgen, wie beispielsweise ein Land oder die ganze Welt zu übernehmen. Dafür würden sie absichtlich komplexe und oft ungeklärte Ereignisse und Phänomene herbeiführen. In Verschwörungstheorien wird von gezielten Täuschungen und Manipulationen ausgegangen. Die geheimen Verschwörer*innen würden jeden Schaden an der Allgemeinheit in Kauf nehmen, um ihr Ziel zu erreichen. Verschwörungsgläubige gehen davon aus, dass die bösen Verschwörer*innen unter Einsatz

betrügerischer Mittel agieren, ganz ohne Rücksicht auf Verluste – weil sie durch und durch böse sind. Die klare Einteilung der Welt in Gut und Böse nennt sich ein manichäisches Weltbild. Die Amadeu Antonio Stiftung stellt dieses Weltbild anhand der Kategorien Selbstbild vs. Feindbild dar (z.B. Selbstbild: das Gute, Feindbild: das Böse; Selbstbild: „wir hier unten", Feindbild: „die da oben"; Selbstbild: das Volk, Feindbild: ein anderes Volk / kein Volk / nicht Teil des Volkes). Menschen mit einem manichäischen Weltbild haben keinen Spielraum für Widersprüche, Unvorhersehbares und Komplexität. Deswegen prallt jede Kritik oder Widerlegung ihrer Behauptungen an ihnen ab.

Eine Verschwörungstheorie besteht in der Regel aus drei Zutaten:

1. den Akteur*innen, also einer geheime, böse Gruppe
2. einer Strategie, die die Gruppe verfolgt
3. dem Ziel der Gruppe

So beispielsweise die Verschwörungserzählung von korrupten Eliten (Akteur*innen), die durch das Einsetzen eines Mikrochips oder durch Impfungen (Strategie) die Kontrolle über die Bevölkerung (Ziel) erlangen wollen. Oder die Erzählung von jüdischen Menschen (Akteur*innen), die Geflüchtete aus anderen Kontinenten nach Europa lenken (Strategie), also Migrations- und Fluchtbewegungen steuern, um die weiße europäische Bevölkerung zu vernichten oder auszutauschen (Ziel) [1].

Verschwörungstheorien in Krisenzeiten

Vor allem in Krisenzeiten greifen Menschen häufiger auf Verschwörungstheorien zurück. Gesellschaftliche Umbrüche, Pandemien oder Kriege führen zu einer großen Verunsicherung in der Gesellschaft. Die Gefahren sind kaum greifbar, was viele Menschen überfordert und ihnen Angst macht. Diese allgemeinen gesellschaftlichen Unsicherheiten fördern Verschwörungstheorien, da diese einen entlastenden Effekt haben. Sie bieten klare Antworten auf das Ungewisse und identifizieren eindeutige „Schuldige" für das Geschehen. Die überwältigende Situation scheint somit strukturiert werden zu können. Die Arbeit der Amadeu Antonio Stiftung bietet eine Analyse der Zusammenhänge zwischen Verschwörungstheorien und Krisen auf sozialer und psychosozialer Ebene. Krisenzeiten, autoritäre Regime, Staatschefs, die Verschwörungsideologien verbreiten, mangelnde Toleranz für Ambiguität, Suche nach Sinn und Klarheit, schwaches und unsicheres Selbstbild sind einige der Kontexte, in denen Verschwörungstheorien wuchern. Einem der bekanntesten Erklärungsansätze zufolge glauben Menschen an Verschwörungstheorien, wenn sie einen Kontrollverlust erleben. Die Kreuzberger Initiative gegen Antisemitismus e.V. spricht außerdem von einer Aufwertung der eigenen Person durch Verschwörungstheorien. Wenn Menschen das Gefühl haben, den wahren Grund oder die wahren Schuldigen hinter unerklärlichen Ereignissen erkannt zu haben, sind sie damit den „Unwissenden" überlegen. Sie bekommen das Gefühl, „das Verborgene durchschaut zu haben und nun zum Kreis der 'Wissenden' zu gehören".

„Wir" gegen „Die Anderen"

Verschwörungstheorien funktionieren nur durch klare Konstruktionen von Eigen- und Fremdgruppen. Dieses „Othering" knüpft an bestehende Ressentiments und Vorurteile an – hierzu gehören zum Beispiel rassistische, antisemitische oder auch sexistische Stereotype. Gängige Feindbilder in Verschwörungstheorien sind z.B. korrupte Politiker*innen, die Lügenpresse, das „Finanzkapital", der Staat bzw. die Regierung, „die Juden" oder „die da oben".

Die Rolle der Sozialen Medien

Das Internet, allen voran die Sozialen Medien, ermöglichen einen schnellen und einfachen Austausch von Menschen, auch wenn sie weit voneinander entfernt und sich noch nie begegnet sind. Diese schnellen und unkomplizierten Möglichkeiten der Vernetzung und der Informationsbeschaffung werden jedoch auch zur Verbreitung von Desinformation und Verschwörungstheorien genutzt. Diese können in Sozialen Netzwerken wie Facebook, Instagram oder TikTok oder Messenger-Diensten wie WhatsApp oder Telegram von den Nutzer*innen geliked und geteilt werden. Dadurch wird die eigene Weltsicht innerhalb der entstandenen Internet-Blase immer wieder bestätigt und die Netzwerke zeigen einem nur noch Inhalte an, die zu den bereits gelikeden und geteilten Inhalten passen. Durch sogenannte Empfehlungsalgorithmen stellen die Betreiber der Plattformen sicher, dass Nutzer*innen das angezeigt bekommen, was sie vermuteterweise sehen wollen. Dadurch kann es zu verzerrten Wahrnehmungen der Realität, zu Filterblasen und Echokammern kommen.

[1] Die rassistische und antisemitische Verschwörungserzählung vom „Großen Austausch" ist übrigens auch fester Bestandteil der Ideologie vieler rechter Attentäter gewesen.

Soziale Medien und Algorithmen

Elisabeth Steffen, Helena Mihaljević und Milena Pustet

Soziale Medien sind aus unserem Alltag nicht mehr wegzudenken: Sie ermöglichen es uns, mit anderen Menschen Informationen zu teilen, Meinungen auszutauschen und soziale Beziehungen aufzubauen und zu pflegen. Über das Internet können wir Fotos, Videos, gesellschaftspolitische Statements und vieles mehr mit der ganzen Welt teilen. Wir halten digital nicht nur Kontakt zu engen Freund*innen, Kolleg*innen oder Verwandten, sondern bauen Beziehungen zu Menschen auf, denen wir sonst vielleicht nie begegnet wären.
Diese Interaktionen finden allerdings nicht auf neutralem Terrain statt, sondern werden geprägt und beeinflusst durch die Algorithmen der Plattformen. In diesem Zusammenhang sind unsere persönlichen Daten und unsere Nutzungsdaten von zentraler Bedeutung für das Geschäftsmodell der meisten Plattformen. Viele Dienste nutzen wir zwar weitgehend kostenlos, „bezahlen" dafür aber mit unseren Daten, welche wiederum in Algorithmen eingespeist werden, um uns beispielsweise personalisierte Werbung anzuzeigen.
Auf den meisten Social Media Plattformen sind solche Empfehlungsalgorithmen allgegenwärtig. Sie beeinflussen dabei nicht nur, welche Werbespots in unserer Timeline erscheinen. Ihre Aufgabe ist es auch, uns auf Basis unseres bisherigen Nutzungsverhaltens sowie den Vorlieben anderer, „ähnlicher" Nutzer*innen Inhalte zu empfehlen. Je mehr Daten der Algorithmus dafür zur Verfügung hat, desto präziser kann er das, was uns vermeintlich gefällt, vorhersagen.

Empfehlungsalgorithmen sind ein mächtiger Filter, der für uns zur Selbstverständlichkeit geworden ist. Sie sorgen dafür, dass uns nur ein extrem begrenzter Ausschnitt der gesamten Interaktion auf der jeweiligen Plattform gezeigt wird. Das Ergebnis sind sogenannte Filterblasen, in denen wir kaum noch mit kontroversen Meinungen oder Ansichten konfrontiert werden, und die wie in einer Echokammer ständig wiederholt werden. Solche Filterblasen können gerade in politischen Diskursen gravierende Effekte haben. Eine Auswertung der Empfehlungen von YouTube während des US-amerikanischen Wahlkampfs verdeutlicht dies besonders gut: Egal ob man auf der Plattform nach Clinton oder Trump gesucht hat - der Algorithmus empfahl mit großer Mehrheit Videos, deren politische Haltung gegen Clinton und pro Trump war, darunter ein signifikanten Anteil von Falschinformationen.[1]

Leider fehlt es an Transparenz seitens der Plattformen, was die genaue Funktionsweise ihrer Algorithmen angeht. Gelegentlich gelangen aber interne Dokumente an die Öffentlichkeit[2], ehemalige Mitarbeiter*innen geben Informationen preis[3], oder Forscher*innen führen Experimente durch, um das Verhalten einzelner Algorithmen zu erforschen und Rückschlüsse über ihre Funktionsweise zu ziehen.[4] Auch wenn jede Plattform ihre eigenen Algorithmen einsetzt, verfolgen sie im Prinzip alle das gleiche Ziel: Die Nutzer*innen sollen möglichst viel Zeit auf der Plattform verbringen. Unterschiede zwischen den Plattformen gibt es eher bezüglich der Frage, welche Nutzer*innen Daten einfließen und mit welcher Gewichtung, und ob zusätzliche Kriterien wie „Account-Verification"[5] miteinfließen.[6] Bei Instagram wird beispielsweise unser Interesse an bestimmten Themen als Hauptfaktor berücksichtigt. Likes und Kommentare sowie die Dauer des Anschauens von Stories und Reels spielen ebenfalls eine Rolle. Zudem wird berücksichtigt, in welcher Beziehung wir zu der Person stehen, die den Content erstellt hat. Der Algorithmus wertet zum Beispiel aus, welche Profile miteinander interagieren, und merkt sich, an wen wir Direktnachrichten verschickt haben.[7]

Während des Präsidentschaftswahlkampfs in den USA im Jahr 2016 zeigte sich auch am Beispiel von Facebook, wie mächtig und zugleich fragil die Rolle von Sozialen Medien und Empfehlungsalgorithmen ist. So geriet Facebook in den Verdacht, eine wichtige Rolle im Zusammenhang mit der Wahl von Trump gespielt zu haben. Die britische Datenanalyse-Firma Cambridge Analytica hatte Zugriff auf Daten von Millionen von Facebook-Nutzer*innen, die sie für personalisierte Werbekampagnen genutzt haben soll. Die Firma soll dabei auch psychologische Profile von Nutzer*innen erstellt haben, um diese gezielt mit politischen Botschaften anzusprechen und so deren Verhalten zu beeinflussen. Bis heute gibt es allerdings keinen eindeutigen Nachweis, dass diese Werbestrategie das Abstimmungsverhalten tatsächlich maßgeblich beeinflusst hat.[8]

Auch Twitter, das zu den relevantesten Plattformen für politische Diskurse zählt, geriet in die Kritik, weil die Empfehlungsalgorithmen dazu beigetragen haben sollen, die Sichtbarkeit von Hasskommentaren und Falschinformationen zu erhöhen. Empfehlungsalgorithmen stehen im Verdacht, polarisierende Inhalte zu befördern - denn in der Regel erhalten zugespitzte Statements, Beleidigungen und auch Hasskommentare mehr Aufmerksamkeit und Reaktionen als ein freundlicher Kommentar oder ein differenziert argumentierender Beitrag.[9] Dabei spielt es keine Rolle, ob unsere Interaktion aus einer zustimmenden oder ablehnenden Haltung heraus stattfindet - was zählt, ist Aktivität an sich.

Ganz grundsätzlich wird am Beispiel Twitter auch deutlich, wie problematisch es sein kann, wenn solche mächtigen Kommunikationsorte kaum öffentlicher Kontrolle unterliegen, sondern vorrangig an den Interessen der Betreiber*innen ausgerichtet sind. Greifbar wurde diese Problematik erst kürzlich wieder im Zusammenhang mit der Übernahme von Twitter durch Elon Musk, der angekündigt hatte, die bisherige Praxis der Content Moderation auf Twitter zugunsten von „free speech" zurückzufahren. In der Folge kam es u.a. zu einem massiven Anstieg antisemitischer Beiträge, deren Anzahl sich nach der Übernahme mehr als verdoppelte.[10] Als Reaktion auf Musks Ankündigungen zogen sich mehrere wichtige Werbekunden von der Plattform zurück, denn wird die Atmosphäre auf einer Plattform zu toxisch, wenden sich viele Nutzer*innen ab, vor allem die von Hassrede Betroffenen.

Deshalb - und auch, weil es in vielen Ländern mittlerweile entsprechende gesetzliche Vorgaben gibt - versuchen die meisten Plattformbetreiber, mit Hilfe von Erkennungsalgorithmen gegen problematische Inhalte vorzugehen. Aufgabe dieser Algorithmen ist es, in der Flut von Postings problematische Inhalte wie z.B. Hass-Posts zu entdecken, um sie dann automatisch oder nach einer zusätzlichen Überprüfung durch menschliche Moderator*innen entfernen zu können. Entsprechende Erkennungsalgorithmen basieren, ähnlich wie Empfehlungsalgorithmen, oft auf Verfahren des maschinellen Lernens, die auf großen Datenmengen trainiert werden. Ausgehend von diesen Trainingsdaten lernen diese Modelle dann auch für neue Beiträge vorherzusagen, ob es sich hierbei um z.B. Hassrede handelt. Durch entsprechende technische Unterstützung können Plattformen schneller Inhalte auswerten, was es grundsätzlich einfacher macht, gegen problematische Inhalte vorzugehen.

Ein Problem stellt allerdings auch hier wieder die Intransparenz bezüglich der Algorithmen dar. In der Regel veröffentlichen Plattformbetreibende keine Details zu den eingesetzten Technologien, wodurch es schwierig ist, die Vorhersagen der Algorithmen nachzuvollziehen und kritisch zu bewerten. Zudem kommt es häufig vor, dass Plattformen erst spät reagieren, wenn Nutzer*innen ihnen problematische Beiträge proaktiv melden. Andere Beiträge hingegen werden gesperrt, ohne dass die genauen Gründe dafür nachvollziehbar sind. Teilweise setzen Plattformen zusätzlich sogenannte Wortfilter ein. Dabei werden problematische Schlüsselwörter und Hashtags in einer Liste gesammelt, und wenn ein*e Nutzer*in einen dieser Begriffe verwendet, wird der Beitrag automatisch entfernt oder blockiert. Diese Methode wurde beispielsweise von TikTok eingesetzt, was jedoch zu Kontroversen führte, da die Plattform auch Begriffe wie „schwul" und „queer" zensierte.[11] Solche Einschränkungen behindern die freie Meinungsäußerung und die Möglichkeit, Aufklärungsarbeit und Community-Empowerment auf Plattformen zu betreiben.

Klar ist jedoch zugleich auch: Ohne automatisierte Verfahren müssten menschliche Mitarbeiter*innen jeden einzelnen Post auf einer Plattform durchgehen, um nach Hass- oder Gewaltinhalten zu suchen. Dies würde nicht nur enorme zeitliche Ressourcen in Anspruch nehmen, sondern auch zu einer starken Belastung für die Mitarbeiter*innen führen, die täglich traumatisierenden Inhalten ausgesetzt wären. Die Verwendung von Erkennungsalgorithmen erscheint daher unerlässlich, um eine effektive Moderation von Social-Media-Plattformen zu gewährleisten. Allerdings basiert die Entwicklung entsprechender Modelle häufig auf der Ausbeutung von Menschen im globalen Süden. Denn tatsächlich bedarf es zunächst einer großen Menge an Trainingsdaten, die von Menschen gelabelt wurden, um ein Modell zu trainieren, das automatisiert toxische Beiträge labeln kann. Um entsprechende Trainingsdatensätze zu erstellen, oder die Ergebnisse der Algorithmen stichprobenartig zu überprüfen, setzen die Plattformen hauptsächlich auf Crowdworking, was größtenteils von Menschen aus ärmeren Ländern, oftmals unter prekären Bedingungen, erfolgt. Auch OpenAI, das Unternehmen hinter ChatGPT, hat die manuelle Überprüfung zahlreicher Texte, die in großer Zahl detaillierte Beschreibungen von Gewalt, Folter oder Selbstverletzung enthielten, an Arbeitende in Kenya outgesourced, mit einer Bezahlung von weniger als 2$ pro Stunde.[12]

Mit der Nutzung von Algorithmus-basierten Moderationsmethoden geht also eine manifeste Ausbeutung von Menschen einher, die oftmals unsichtbar bleibt. Ein weiterer problematischer Aspekt beim Einsatz von automatisierten Verfahren resultiert aus dem Umstand, dass es unterschiedliche Regularien je nach Land gibt und dass sich auch die vorhandenen Datensätze für verschiedene Sprachen in Qualität und Umfang massiv unterscheiden. Dadurch entstehen häufig große Unterschiede in der Qualität der Verfahren, die in den verschiedenen Ländern und Sprachräumen zum Einsatz kommen. Auch dadurch werden Hierarchisierungen und Machtverhältnisse zwischen dem globalen Süden und Norden reproduziert und verfestigt.

Und schließlich spiegeln Trainingsdaten oft gesellschaftliche Machtverhältnisse wider, die dann von den Modellen erlernt und reproduziert werden. So stellte sich beispielsweise heraus, dass Amazon's Algorithmus für Neueinstellungen, trotz intensiver Bemühungen seitens Amazon das Modell zu „debiasen", einen gender bias aufwies und Männer insbesondere für IT Berufe bevorzugte.[13] Noch drastischer sind die Konsequenzen, wenn Algorithmen bei Gerichtsprozessen eingesetzt werden. In den USA zeigte sich dabei, dass schwarze Menschen gegenüber weißen Menschen deutlich benachteiligt wurden – für sie wurde fast doppelt so oft falsch vorhergesagt, dass sie rückfällig werden würden.[14]

All dies macht deutlich, welche Herausforderungen mit dem Einsatz von automatisierten Verfahren zur Content Moderation in Sozialen Medien einhergehen.

Auf dem Symposium „Digitaler Hass" im September 2022 in Berlin wurden diese Herausforderungen in einem eigenen Panel adressiert.[15] Dabei stellte Leah Nann das Projekt AI4Dignity vor, das sich mit einem bottom-up Ansatz des „Ethical Scaling" für kollaborativ und inklusiv entwickelte Trainingsdatensätze und Algorithmen für Content Moderation einsetzt. Helena Mihaljević ging anhand des Tools Perspective API der Frage nach, ob ein Modell zur Erkennung toxischer Sprache geeignet ist, antisemitische Inhalte zu erkennen. Anne Kaun beschäftigte sich mit der menschlichen Arbeit hinter den Algorithmen, die unter anderem auch in Form von Gefängnisarbeit geleistet wird.

Dabei wurde zugleich immer wieder deutlich: Algorithmen sind längst Teil unseres Alltags und werden aus unserer Gesellschaft nicht mehr verschwinden. Deshalb ist es umso wichtiger, einen verantwortungsvollen Umgang mit ihnen zu finden. Wissenschaftliche Forschung kann hierbei einen wichtigen Beitrag leisten: Sie kann bestehende Algorithmen untersuchen und deren Schwachstellen und Verzerrungen offenlegen. Sie kann aktiv den Austausch mit der Zivilgesellschaft suchen und fördern, um herauszufinden, welche Probleme durch Algorithmen entstehen und wie diesen in der gesellschaftlichen Realität begegnet werden kann. Sie kann auch dazu beitragen, eigene Datensätze und Modelle für die Klassifikation problematischer Inhalte zu entwickeln, um transparente und differenzierte Verfahren für die Moderation von Inhalten bereitzustellen. Sie kann Prozesse und Kriterien für die Evaluation von Modellen kritisch hinterfragen und Diversität und Inklusion in diesem Kontext stärken, etwa bei der Festlegung von Evaluationskriterien oder bei der Zusammensetzung von Entwickler*innen-Teams. Und auch die globalen Produktionsbedingungen von Algorithmen müssen durch sie kritisch hinterfragt werden, um eine Ausbeutung des globalen Südens bei der Erstellung von Datensätzen zu verhindern. All dies kann dazu beitragen. dass Algorithmen in Zukunft besser eingesetzt und vor allem besser hinsichtlich ihrer gesellschaftlichen Auswirkungen evaluiert werden.

[1] https://medium.com/the-graph/youtubes-ai-is-neutral-towards-clicks-but-is-biased-towards-people-and-ideas-3a2f643dea9a

[2] https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html

[3] Guillaume Chaslot, ein ehemaliger Software Entwickler bei Google, dem Konzern, zu dem auch YouTube gehört, hat beispielsweise eine Software entwickelt, mit welcher man die Empfehlungen von YouTube zu einer beliebigen Suchanfrage auswerten konnte. https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot

[4] https://algorithmwatch.org/de/risikofalle-social-media/ https://www.rnd.de/digital/algorithmen-wenn-maschinen-auf-social-media-plattformen-entscheiden-HIS5N3SXRVE3THWPVFILKIYSXE.html

[5] https://www.cnbc.com/2023/03/28/elon-musk-says-only-verified-twitter-users-to-show-up-in-for-you-tab.html

[6] Gute Übersicht mit Quellen zu versch. Plattformen https://www.ndr.de/ratgeber/medienkompetenz/Wie-wirken-Algorithmen-Unterrichtsmaterial-fuer-die-Schule,algorithmus100.html

[7] https://www.ndr.de/ratgeber/medienkompetenz/algorithmen100.pdf

[8] https://www.nature.com/articles/d41586-018-03880-4 https://www.wired.com/story/viral-political-ads-not-as-persuasive-as-you-think/

[9] So zeigen Studien, dass Nutzer*innen eine höhere Bereitschaft zur Interaktion gegenüber negativen oder kontroversen Inhalten zeigen. https://journals.sagepub.com/doi/pdf/10.1177/1461444813495165, https://www.researchgate.net/publication/281215058_Der_virtuelle_Stammtisch_Determinanten_interpersonal-offentlicher_Kommunikation_auf_Nachrichtenwebsites

[10] https://theconversation.com/antisemitism-on-twitter-has-more-than-doubled-since-elon-musk-took-over-the-platform-new-research-201830

[11] https://www.tagesschau.de/investigativ/ndr/tik-tok-begriffe-101.html https://www.tagesschau.de/investigativ/tik-tok-begriffe-blockade-101.html

[12] https://time.com/6247678/openai-chatgpt-kenya-workers/

[13] https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

[14] https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[15] https://archiv.hkw.de/de/app/mediathek/video/96194

Rassismus-kritische Überlegungen in digitalen Zeiten

María do Mar Castro Varela

Ein Blick auf die Möglichkeiten von Algorithmen führt uns auf die Spuren digitaler rassistischer Praxen. Sie lassen uns auch die Heterogenität historischer Gewaltgeschichten sehen und vielleicht auch besser verstehen. Jegliche Untersuchung von Hass-Reden geschieht in einem spezifischen Kontext, d.h. Zeit und Raum. Die digitale Welt, die bereits vor der COVID-19 Pandemie unser Leben bestimmt hat, tut es seitdem noch effektiver. Zahlreiche Studien beschäftigen sich heute mit sogenannten „online-offline-Spillovers“, d.h. sie untersuchen, wie die Kommunikation auf digitalen Plattformen das soziale Verhalten offline bestimmt. Die Studien kommen, je nachdem wie Spillovers analysiert und definiert betrachtet werden, selbstredend zu unterschiedlichen Ergebnissen. Doch alle scheinen sich einig: Spillovers sind eine Tatsache. Bekannt ist etwa das hasserfüllte digitale Verhalten von den Tätern in Halle oder Hanau: „Das Internet hat für rechtsextreme Organisationen und rassistische und antisemitische Gewalttäter in den vergangenen beiden Jahrzehnten massiv an Bedeutung gewonnen: Insbesondere soziale Medien und Videoplattformen spielen eine zentrale Rolle für die Verbreitung von Propaganda und die Rekrutierung von Nachahmern.“ (Bundeszentrale für politische Bildung 2020, o.S.) Interessant ist hier aber auch die zunehmende Verwischung der Grenze zwischen „digital“ und „real“. Die digitale Welt, so könnten wir sagen, ist lange schon real.

Fragen, die sich daran anschließen, sind etwa folgende: Wie entsteht Wissen im digitalen Zeitalter? Wie werden diskursive Räume aufgebaut und verteidigt? Wie wird Hass kultiviert, abgewehrt und rekalibriert? Wie werden analoge Verschwörungstheorien digitalisiert? Einige Antworten darauf gibt es bereits, doch müssen die Fragen immer wieder neu gestellt werden, denn die Welt der sozialen Medien wird immer bunter und undurchdringlicher. Längst schon sprechen wir nicht nur von Tweets und Memes, sondern auch von der Macht der Algorithmen und wie diese manchmal unbemerkt unsere Vorlieben steuern und auch unser Wissen kontrollieren. Künstliche Intelligenz ist zum „großen Thema“ geworden: verteufelt und gefeiert. Ein guter pädagogischer Umgang damit muss noch gefunden werden. Nach einer ersten hysterischen Reaktion auf ChatGPT etwa wird nun versucht, dieses im pädagogischen Alltag zu nutzen, während gleichzeitig Ethikkommissionen über die Kontrolle von künstlicher Intelligenz nachdenken. Manche rufen nach einem Entwicklungsstopp, weil die rasanten Entwicklungen mit unseren ethischen Handlungsmöglichkeiten kollidieren.

Wer Hass-Reden heute verstehen will, muss deswegen auch die materiellen Möglichkeiten und die digital-politischen Verhältnisse mit in den Blick nehmen. Hass wird heute in Form von „cut-and-paste“ disseminiert – und zwar transnational. Rassistische Memes können in einem deutschen Account auftauchen und in der nächsten Sekunde in Taiwan. Für ein pädagogisches Nachdenken bedeutet dies, dass zum einen mehr als zuvor provinzielle Lösungen ins Leere laufen. Zum anderen wird es immer wichtiger, dass Menschen, die in der Vermittlung tätig sind, über ein immer größeres Know-how bezüglich der digitalen Bedingungen verfügen müssen.

Hass-Archive und rassismuskritische Literalität

Suchmaschinen und Algorithmen produzieren und verarbeiten in Sekundenschnelle unvorstellbare Mengen

an Daten. Diese wiederum speisen riesige Archive, die beständig wachsen. Alles was wir schreiben, die Bilder, die wir hochladen, aber auch Websites, Homepages, Tweets and Videos sind in einem unendlichen Archiv, wo sie nicht stillstehen, sondern kursieren und ständig aufgerufen, verworfen und organisiert werden.

In seiner Schrift *Dem Archiv verschrieben* (1997) räsoniert der Philosoph Jacques Derrida über die Obsession des Erstellens von Listen, über die Suche nach Erinnerung und diese unglaubliche Wut, nicht vergessen zu wollen. In rassismuskritischen Kontexten wird der Erinnerung häufig eine durchweg positive Rolle zugewiesen, während Verdrängung und Vergessen als Problem oder das Böse beschrieben werden. Doch die Obsession des Archivierens, dieses Fieber, mit dem gesammelt, katalogisiert und systematisiert wird, verweist bei Derrida auch auf eine Kehrseite: Archive beherbergen Schriften, Ideen, Imaginationen, Bilder, aber auch Affekte und Gefühle. Objekte und Gegenstände, die in Archiven aufbewahrt, ja, gefangen gehalten werden, können jederzeit kursieren, also in Bewegung geraten und dadurch informieren, aber auch affizieren. Derrida bemerkt, dass „die Archivierung [...] das Ergebnis in gleichem Maße" hervorbringe, wie es diese aufzeichne (Derrida 1997, S. 35). Archive sind keine passiven Aufbewahrungsorte. Die Macht der Katalogisierung sowie die Kontrolle des Zugangs machen sie zu einer produktiven Entität (siehe auch Castro Varela/Shure 2021).

Übertragen wir die Ideen, die wir in Dem Archiv verschrieben finden, auf den Bereich der Datenarchivierung, so stoßen wir rasch auf die ethischen Arbeiten von Mutale Nkonde. Nkonde ist eine afro-amerikanische KI-Forscherin, die früh schon auf den rassistischen Bias von Algorithmen hingewiesen hat. Dabei zeigt sie auf, dass auch die Mathematik nicht frei von Ideologie ist bzw., dass sie für ideologische Zwecke instrumentalisiert werden kann. Nkonde führt etwa aus, dass Algorithmen immer im Zusammenhang mit jenen Menschen stehen, die sie programmieren (siehe auch Peteranderl 2019). Dies ist bekannt, doch plädiert sie eben aus diesem Grunde für eine rassismuskritische Literalität im IT-Bereich (Daniels/Nkonde/Mir 2019). Nur wenn die Programmier:innen (die immer noch weit überwiegend weiß, männlich, heterosexuell und cis sind) ihre rassistischen Vorurteile kennen und überwinden, werden die Algorithmen selber weniger rassistisch sein (siehe Bernabeu 2017). Ein vertieftes Wissen über die Geschichte, Praxen und Wirksamkeit von Rassismus, eine Bildung im Feld des Rassismus ist deswegen auch für jene unabdingbar, die Programmierungen durchführen oder im Bereich der KI arbeiten. Gleichzeitig bleibt eine politische Bildung notwendig, die Schüler:innen und Studierende darüber informiert, wie Algorithmen funktionieren, wie diese entlarvt und ein Umgang mit ihnen gefunden werden kann. Es geht gewissermaßen um eine Ideologiekritik 3.0.

In *The Costs of Connection* zeigen Nick Couldry und Ulises Mejias (2019), welche Auswirkungen die zunehmende Datafizierung auf die Gesellschaft, die Demokratie sowie auf das menschliche Leben im Allgemeinen hat. Die Autor:innen bezeichnen den gegenwärtigen sozialen Zustand als „Datenkolonialismus". Damit weisen sie auf eine neue und gefährliche Verstrickung zwischen Kolonialismus und Kapitalismus hin. Verstehen wir Kolonialismus als einen komplexen geopolitischen Prozess, bei dem eine Gruppe von Menschen unter Einsatz von Technologien der Überwachung und Kontrolle den Lebensraum anderer Menschen besetzt, deren Ressourcen ausbeutet sowie die eigenen Ideologien durchsetzt, so haben wir es mit einer neuen Ära des Kolonialismus zu tun (Couldry/Mejias 2019, S. 49). Die großen Technologieunternehmen verstehen Daten in dieser neuen Ära als eine begehrenswerte Ware, die gesammelt, angeboten, gestohlen und verkauft werden kann. Für Couldry und Meijas haben wir es mit einer Kontinuität des Kolonialismus zu tun, weil die gleiche Logik am Werk ist. Den kolonialen Mächten ermöglichte beispielsweise die Definierung der kolonisierten Länder als „leere Räume", die Bodenschätze, die Natur, die kulturellen Artefakte sowie die Menschen unbeschränkt auszubeuten und anzueignen (siehe Castro Varela/Dhawan 2020, S. 23).

Digitale Technologien ermöglichen heute die Archivierungen von Daten, Texten, Bildern, Objekten etc. in ganz neuen Dimensionen. Archiviert werden etwa auch die Gewohnheiten und Vorlieben von Nutzer:innen. Algorithmen lernen von den Präferenzen der Nutzer:innen im Netz und können so künftige Handlungen, Einstellungen und Entscheidungen nicht nur prognostizieren, sondern auch steuern. Äußert beispielsweise eine Person im Internet Bedenken gegenüber einer Impfung gegen COVID-19, da sie gehört hat, Bill Gates wolle mit dem Impfstoff die DNA von Menschen verändern, so ist die Wahrscheinlichkeit recht hoch, dass diese Person bei der nächsten Verwendung einer Suchmaschine mit Angeboten ähnlicher Verschwörungserzählungen regelrecht überflutet wird. Entsprechende Narrative und Bilder tauchen dann nicht nur auf einer Website oder auf einem Gerät auf, sondern diese schmuggeln sich durch alle von der Person genutzten digitalen Geräte: das Smartphone, das Tablet, die Newsfeeds auf dem Laptop etc. Das komplexe Netzwerk unterschiedlicher digitaler Geräte ermöglicht es dabei nicht nur Unternehmen, sondern auch Regierungen sowie anderen politischen Akteur:innen, Daten von Menschen zu sammeln und deren politische Handlungen zu steuern.

Algorithmen unterstützen so auch die rasche und massive Verbreitung von Desinformationen. Darüber hinaus greifen sie auch in die Entscheidungsfähigkeit der Internetnutzer:innen ein. Die COVID-19 Pandemie hat etwa Millionen von Menschen weltweit dazu genötigt, Waren über das Internet zu bestellen. Die großen Gewinner der Pandemie sind daher Online-Warenhäuser wie Amazon, die ihre Gewinne um ein Vielfaches erhöhen konnten. Mussten wir früher riesige Werbeplakate ertragen und uns eventuell darüber ärgern, dass die Fenster im Bus mit Werbung überklebt waren, so begleitet uns eine auf uns zugeschnittene Werbung nun tagtäglich, ja minütlich auf unseren digitalen Endgeräten. Werbeblocker halten schon lange nicht mehr, was sie versprechen.

Was sich beim Marketing von Waren als nützlich erwiesen hat, wird seit Langem schon auch für die Manipulation politischer Entscheidungen genutzt. Durch die eine gezielte ideologische Ausrichtung und die ständige

Wiederholung von Inhalten werden Nutzer:innen von Sozialen Medien in einer Art Dauerschleife gefangen. Jede Wiederholung trägt dazu bei, dass sie das, was sie anfangs gelesen haben, noch mehr glauben. „Wissen“ wird über eine ständige Wiederholung von „Glaubenssätzen“ durchgesetzt. Die Vielfalt von Perspektiven wird eingeengt auf einige wenige Ideen, Aussagen und Ansichten. Das hat allerdings – wie Studien zeigen – auch damit zu tun, dass in Europa schon seit einigen Jahrzehnten immer mehr Menschen dazu tendieren, sich nur mit möglichst ähnlichen Mitmenschen zu umgeben. Das beeinflusst den Wohnort, den wir wählen, ebenso wie den Sport, den wir treiben, oder die Filme, die wir auf YouTube sehen (siehe Bishop 2009). In der Soziologie ist in diesem Zusammenhang die Rede von sozialer Homophilie (McPherson et al. 2001, S. 416). Digitale Netzwerke leben geradezu von einer sozialen Homophilie. Wenn zusätzlich noch Algorithmen dafür sorgen, dass wir auf Facebook, Twitter oder YouTube immer wieder Beiträge, Tweets oder Feeds angeboten bekommen, denen wir tendenziell zustimmen, bewegen wir uns in einer Informationsblase. Diese kann sich zunächst angenehm anfühlen. Wir bekommen Informationen, die uns interessieren, präsentiert in einem ideologischen Vokabular und Stil, die uns glauben machen: „Das ist richtig!“ Wir erhalten Anerkennung und Bestätigung für unsere eigene Meinung und empfinden das als positiv. Dazu gesellen sich Memes, Sticker, Emojis oder kurze Videos, die die Mitglieder einer Blase choreografiert zum Lachen bringen oder wütend machen können. Aus demokratiepolitischer Sicht ist dies mehr als bedenklich. Eingeübt wird ein exklusives Denken, das es uns immer schwerer macht, andere Vorstellungen auch nur wahrzunehmen – geschweige denn zu akzeptieren.

Dennoch oder gerade deshalb ist es wichtig, eine unregulierte Weiterentwicklung von KI zu verhindern, denn diese bestimmt insbesondere die Situation marginalisierter Menschen. Safiya Umoja Noble (2018) zeigt sich beunruhigt darüber, dass Maschinen Entscheidungen darüber treffen, wer Leben darf, wer kriminell ist, wer Kredite aufnehmen darf, wer überwacht werden soll etc. Noble glaubt, dass Algorithmen und insbesondere die KI zentrale Themen sind, die wir in nicht mehr aus den Augen verlieren dürfen, insbesondere, wenn es um Fragen der Menschlichkeit geht (ebd., S. 28). Dem stimmen auch Damini Gupta und T. S. Krishnan (2020) zu. Allerdings richten sie unseren Blick auf Kämpfe, die bereits geführt werden. Die beiden sind weniger pessimistisch und nennen verschiedene Kollektive, die sich im Kampf gegen stereotype KI-Algorithmen gebildet haben. Sie bemerken auch, dass Tech-Unternehmen es sich schon bald nicht mehr werden leisten können, rassistische Algorithmen zu produzieren. „Der Wind der Veränderung weht von sozialen Aktivisten, die sich mit den geschlechtsspezifischen, rassistischen und sozioökonomischen Vorurteilen von KI-Algorithmen auseinandersetzen und die Regierungen dazu drängen, Gesetze zu entwickeln, die diese Algorithmen regeln. Ihre Bemühungen erhöhen auch die Kosten von Reputationsverlusten für Unternehmen, indem sie ein öffentliches Bewusstsein für die Auswirkungen von voreingenommenen KI-Algorithmen schaffen.“ (Gupta/ Krishnan 2020, o.S.)

Demokratie in digitalen Zeiten

Auch in der Politik wird immer häufiger über die Gefahren der rasanten Datafizierung für Demokratien gesprochen. Am 23. Oktober 2019 konnten wir beispielsweise eine historische Befragung des Facebook-Gründers Mark Zuckerberg durch die Demokratische US-Senatorin Alexandria Ocasio-Cortez mitverfolgen. Gegenstand war der Datenskandal um Facebook und Cambridge Analytica. Letztere hatte sich Zugang zu mindestens 80 Millionen Facebook-Profilen verschafft und deren Daten ausgewertet, um das Wahlverhalten der Nutzer:innen zu manipulieren. Cambridge Analytica nutzte die Daten etwa, um die Präsidentschaftskampagnen von Ted Cruz und Donald Trump 2016 zu unterstützen. Facebook musste sich schließlich für seine Rolle bei der Datenerfassung entschuldigen. Sichtbar wurde aber, wie fragil unsere Demokratien sind.

Sicher wird ein Workshop für Lehrer:innen oder ein Seminar für Studierende diese Gefahren nicht bremsen können, doch scheint es wichtig, über gute digitale pädagogische Interventionsmöglichkeiten nachzudenken. Denn wenn es auch gut ist, dass einige Arbeiten bald schon von Robotern oder gesteuert durch KI durchgeführt werden (etwa Botengänge oder die Müllabfuhr), ist die Gefahr der Manipulation und damit die Entdemokratisierung der Bürger:innen sehr ernst zu nehmen.

Eine politische Bildung heute muss deswegen nicht nur darlegen, wie rassistischer Hass digital verbreitet wird, wer davon profitiert und wie online Hass-Reden zu offline Hasstaaten führen. Eine zeitgemäße politische Bildung muss auch dafür sorgen, dass nach Strategien gesucht wird, die eine politische Manipulation erschwert. Dazu bedarf es einer digitalen ebenso wie einer rassismuskritischen Literalität.

Literatur

Bishop, Bill (2009): The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart, Boston: Houghton Mifflin Harcourt.

Bundeszentrale für politische Bildung (2020): »Der Anschlag von Halle«, in: https://www.bpb.de/kurz-knapp/hintergrund-aktuell/316638/der-anschlag-von-halle/ (01.11.2023).

Castro Varela, María do Mar/Dhawan, Nikita (2020): Postkoloniale Theorie. Eine kritische Einführung, (3. überarbeitete und ergänzte Auflage), Bielefeld: transcript (UTB).

Castro Varela, María do Mar/Shure, Saphira (2021): »Archiv-Fieber: Kritische Erinnerungsarbeit in den Bildungswissenschaften“, in: Vierteljahreszeitschrift für wissenschaftliche Pädagogik H. 3/2021, 97. Jahrgang, S. 286-302.

Couldry, Nick/Mejias, Ulises A. (2019): The Costs Of Connection: How Data Is Colonizing Human Life And Appropriating It For Capitalism, Stanford: Stanford University Press.

Daniels, Jessie/Nkonde, Mutale/Mir, Darakhshan (2019): Advancing Racial Literacy In Tech. Why Ethics, Diversity in Hiring & Implicit Bias Trainings Aren’t Enough. Data & Society, https://datasociety.net/wp-content/uploads/2019/05/Racial_Literacy_Tech_Final_0522.pdf (01.11.2021).

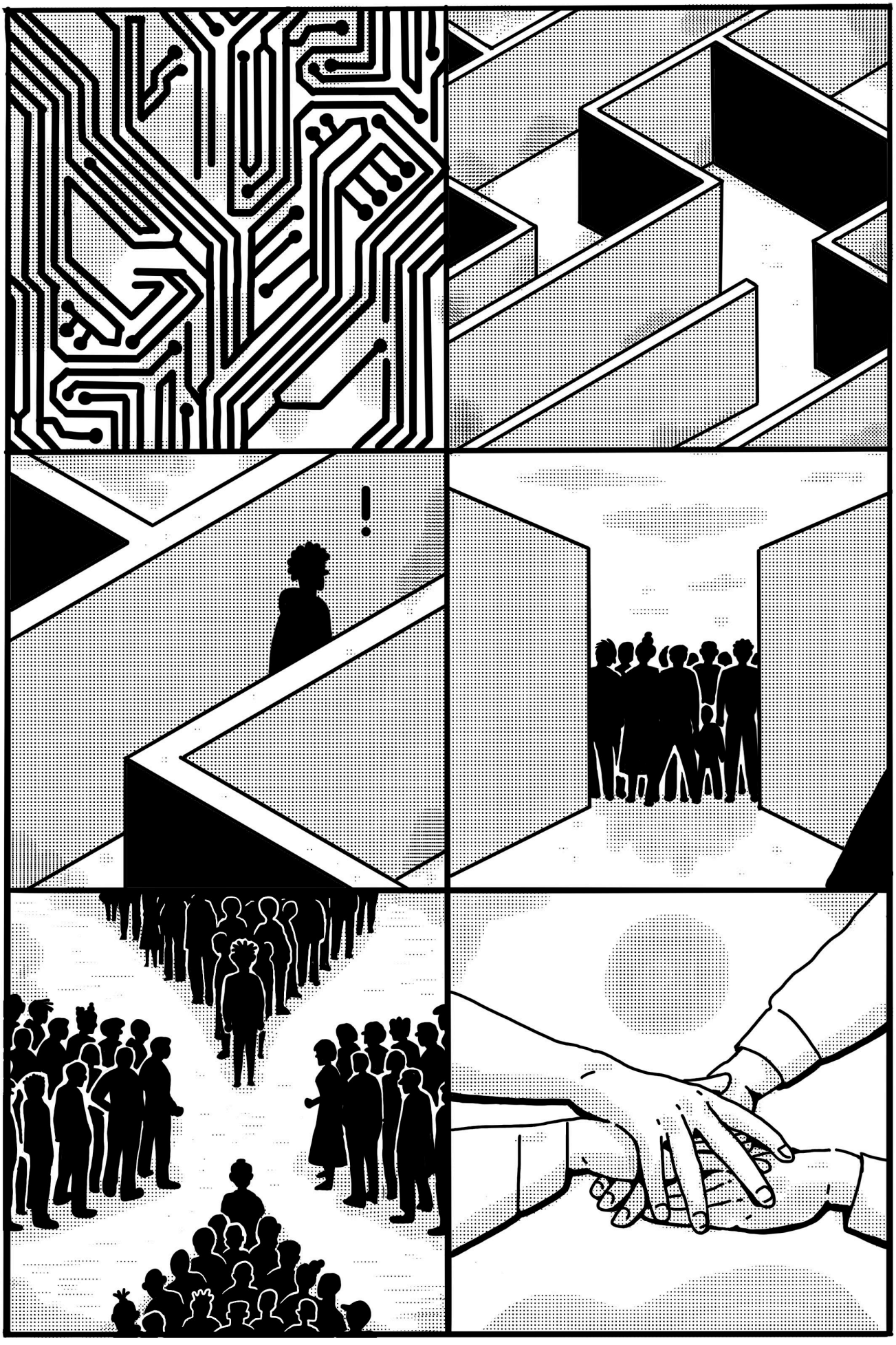
Derrida, Jacques (1997): Dem Archiv verschrieben, Berlin: Brinkmann + Bose.

Gupta, Damini/Krishnan, T. S. (2020): »Algorithmic Bias: Why Bother?«, in: California Management Review 63 (3), https://cmr.berkeley.edu/2020/11/algorithmic-bias/ (01.11.2023).

McPherson, Miller et al. (2001): »Birds of A Feather: Homophily In Social Networks«, in: Annual Review of Sociology 27, S. 415-444.

Noble, Safiya Umoja (2018): Algorithms of Oppression. How Search Engines Reinforce Racism, New York: New York University Press.

Peteranderl, Sonja (2019): »Wie viel Rassismus steckt in Algorithmen?«, in: Der Spiegel vom 15.06.2019, https://www.spiegel.de/netzwelt/netzpolitik/rassistische-algorithmen-ki-forscherin-mutale-nkondeim-interview-a-1271778.html (01.11.2023).



Gegen interpersonelle und algorithmische Formen von Gewalt

Konflikte und Widersprüche in der Architektur des digitalen Raums

Verónica Orsi

Dieser Text entstand aus einem Gespräch mit Marcus Roeper. Er arbeitet bei Birds on Mars, ist KI-Experte und gibt Beratung, Lectures und Workshops zu den Themen Digitaler Hass und Künstliche Intelligenz. Außerdem ist er Developer und experimentiert mit Modellen, die Hass generieren, um durch Hands-On-Ansätze Bewusstsein für die Gefahren von verbaler Gewalt zu schaffen. Wir haben uns über Hassreden und die Verhaltensnormen im digitalen Raum sowie über Künstliche Intelligenz und die ethischen und moralischen Implikationen, eine Maschine zu trainieren, unterhalten.

Mit der rapiden Entwicklung der künstlichen Intelligenz betreten wir ein Szenario, das schwer zu beschreiben ist. Unser Dasein in der Welt verändert sich angesichts eines Problems, das wir noch nicht bewältigen können, weil seine Folgen unsere Vorstellungskraft übersteigen. Unsere zwischenmenschlichen Beziehungen und die Beziehungen zu unserer Umwelt sind zunehmend mit algorithmischen Architekturen und intelligenten Modellen verwoben, die die Art und Weise, wie wir die Welt sehen und mit ihr interagieren, bestimmen. Die Form der Menschheit, wie wir sie kennen, wandelt sich und wir verharren in einem Zustand der digital-analphabetischen Träumerei. In der digitalen Welt wie auch in der realen Welt verschieben sich die zwischenmenschlichen Verhaltensnormen in eine Richtung, aus der es schwierig wird, wieder herauszukommen. Die digitale Anonymität ist für Marcus Roeper eine der Grundlagen dieser Verschiebungen und grundlegend für die Ausbreitung von Verantwortungslosigkeit und Gewalt im virtuellen Raum. Wir sind mit neuen Formen missbräuchlicher Beziehungen, neuen Formen der Diskriminierung und des Hasses, neuen Formen der Unterdrückung in Bezug auf Repräsentation, Sichtbarkeit und politische Teilhabe konfrontiert, und wir alle haben dabei den Eindruck, mit diesem großen kollektiven Problem allein zu sein.

Es gibt eine Reihe von Themen, die transdisziplinär und kollektiv, mit intellektuellem Mut und Empathie angegangen werden müssen, um aus diesem Labyrinth herauszufinden. Einem Labyrinth, das wir selbst geschaffen haben, das wir aber weder beschreiben noch zerstören können. Zum einen ist die Architektur des digitalen Raums nicht für den Menschen gemacht. Das bedeutet, dass die Formen der Inhalte und die Instrumente der Interaktion mit anderen Nutzern darauf ausgerichtet sind, die Maschine zu bereichern und nicht uns als sensible, denkende und fühlende Wesen. Der anonyme Hass zwischen Menschen, der von Maschinen geschürt

wird, die ethischen Widersprüche in der Art und Weise, wie künstliche Intelligenz trainiert wird und die Gefahren von Fakes und Misleadings zwingen uns, über eine Kombination von Strategien aus Sozialpädagogik, Diversity-Arbeit, staatlicher Politik, Privatwirtschaft und affektivem Training von Maschinen nachzudenken, um aus einer vielleicht fatalen Zukunft herauszukommen und sich den neuen Herausforderungen zu stellen, die die Entwicklung der Technologie mit sich bringt.

Hass
Das Unpersönliche der Verhaltensnormen im digitalen Raum

Ausgehend von der Prämisse, dass eine Welt ohne digitale Interaktion zukünftig unvorstellbar ist, ist es unerlässlich, darüber nachzudenken, wie soziale Netzwerke und künstliche Intelligenz funktionieren. Beginnen wir mit den sozialen Netzwerken. Im Prinzip funktionieren soziale Netzwerke, weil wir alle dort sind. Wenn die sozialen Netze keine Nutzer*innen mehr haben, bleiben die digitalen Fußabdrücke des Netzwerks bestehen, aber ihre Aktivität stirbt. Die Interaktion zwischen den Nutzer*innen eines populären Netzwerks wird von den Firmen, die es betreiben, durch den Verkauf von Werbung und die Extraktion und Analyse des Nutzungsverhaltens kapitalisiert. Jede Minute Aufmerksamkeit, die wir den sozialen Netzwerken schenken, ist eine weitere Minute, in der wir Werbeanzeigen konsumieren und Informationen über unser Verhalten an die Unternehmen weitergeben. All dies geschieht, während wir mit unseren Freund*innen in Kontakt treten, uns Fotos und Videos ansehen oder Informationen abrufen. Die Gefahr dieses Geschäftsmodells besteht darin, dass die Unternehmen, denen die sozialen Netzwerke gehören, unsere Aufmerksamkeit und möglichst viele Minuten unseres ‘Engagements’ brauchen, um mehr Werbung verkaufen zu können. Um dies zu erreichen, bieten sie uns in unserem Feed Inhalte an, die sicherstellen, dass wir online bleiben, und die algorithmisch speziell auf uns zugeschnitten sind. Statistisch gesehen interagieren Menschen mehr mit Inhalten, die sie wütend und traurig machen, als mit Inhalten, die sie erfreuen, so Marcus Roeper. Deshalb besteht die sicherste Grundlage dieses Geschäftsmodells darin, die menschliche Affektivität auszunutzen, umso für mehr und anhaltenden Umsatz zu sorgen.

Die Frage des Hasses in sozialen Netzwerken und im digitalen Raum ist ein delikates und komplexes Thema, das in seiner Multidimensionalität analysiert werden muss. Wenn es um gewalttätige Interaktionen zwischen Nutzer*innen auf verschiedenen digitalen Plattformen geht, ist nicht zuletzt auch die Anonymität des Raums eine notwendige Grundlage für den Angriff. So ist das digitale Verhalten aufgrund des Designs der Plattformen unfreundlicher, hässlicher und fast asozialer geworden. Das Internet war nie ein angenehmer und hassfreier Raum. Es ist nur einfacher geworden, eine Plattform mit schlechtem Verhalten und bösartigen Meinungen zu finden, ist Marcus Roeper überzeugt. Zugleich schlägt sich der Hass in den Netzwerken auch in realem und konkretem Hass, Diskriminierung und Rassismus in der nicht-virtuellen Welt nieder. Während also eindeutig rassistische Beleidigungen auf Twitter, Facebook und Instagram seltener zu finden sind, sind andere verschlüsselte Formen des Hasses sichtbar, und dies wird auch in der realen Welt wahrgenommen. Andere Plattformen mit weniger regulierten Nutzungsrichtlinien enthalten offen hasserfüllte Kommentare, so dass es schwierig ist, an eine einzige Interventionsmaßnahme zu denken.

Marcus Roeper arbeitet unter anderem als KI-Experte bei die Deutsche Bahn (DB) und berät bei der Verarbeitung von Daten, die durch das Feedback der Zugnutzer*innen entstehen. Die DB hat ein Feedback-System, bei dem Fahrgäste das Recht haben, ihre Meinung zu äußern oder den Service an Bord zu beurteilen. Auf dieser Plattform können unzufriedene Fahrgäste ihre Kritik direkt gegenüber dem Zugpersonal äußern, ohne dieses persönlich adressieren zu müssen. Dies führt dazu, dass viele Menschen, die mit dem Service unzufrieden sind, die unter Verspätungen und Zugausfällen gelitten haben und die Zeit und Geld verloren haben, aufgrund der Gestaltung der digitalen Plattform die Mitarbeiter*innen bösartig angreifen können. Wenn diese Kommentare von den Zugbegleitern direkt gelesen würden, hätte dies schwerwiegende Folgen wie Angst, Stress und Leid. Diese zum Teil unglaublich harten, absichtlich bösartigen Nachrichten kommen ungefiltert bei den Mitarbeitenden an, die nichts anderes tun können, als sich dem zu stellen und darunter zu leiden. Der KI-Entwickler ist zwiegespalten, wie mit dieser Situation umgegangen werden soll. Einerseits könnte man darüber nachdenken, eine Maschine so zu trainieren, dass sie aggressive Nachrichten herausfiltert. Diese Lösung würde jedoch bedeuten, dass nur positive Nachrichten von Fahrgästen, die glücklich und zufrieden mit dem Service sind, gespeichert werden. Eine andere Lösung wäre, die gewaltvolle Mitteilung mit einer neutralen Mitteilung zu überdecken, bei der lediglich die darin enthaltene Information übermittelt wird, die gewalttätige Aussage jedoch nicht.

Marcus Roeper ist überzeugt, dass für diese relativ neuen Formen der Interaktion (sowohl in sozialen Netzwerken als auch in anderen digitalen Räumen) bald ein ganzes Netz an Lösungen gefunden werden muss, da die Anonymität dieser Plattformen zu grausamen Verhaltensweisen und unbekannten Formen von Gewalt führt, die sowohl online als auch offline Auswirkungen haben. Er ist der Ansicht, dass sich die Menschen im Kontext der sozialen Medien immer in Hassspiralen verlieren werden, die dafür geschaffen und gestaltet wurden, da es für bestimmte Gruppen einen ganz besonderen Wert hat, wenn Menschen in den sozialen Medien mit Hass interagieren. Im Kontext anderer Interaktionsplattformen, die auf Anonymität basieren, gibt es andere Barrieren und Herausforderungen, denen es sich zu stellen gilt, wenn nach Lösungen für ein humaneres, bescheideneres und empathischeres Miteinander gesucht wird.

Künstliche Intelligenz
Die ethischen und moralischen Grenzen, eine Maschine zu trainieren

Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz, bei dem eine Maschine automatisch eine Masse von Daten zu einem bestimmten Thema lernt, um diese Informationen zu verarbeiten. Was eine öffentliche

Debatte unter Experten, Wissenschaftlern, politischen Vertretern und inzwischen auch der Zivilgesellschaft auslöst, ist, dass nicht klar ist, wie die Maschinen gefüttert werden, wer sie füttert und was die möglichen Folgen dieser Art von Technologie sind, da es noch keine Präzedenzfälle gibt. Bei der Entwicklung dieser Technologie herrscht im Umgang ein riesiger Mangel an Transparenz über die Prozesse, die hinter den Kulissen ablaufen. Das Ziel ist klar: Die künstliche Intelligenz soll uns so effizient wie möglich dienen und dabei so wenig Zerstörung wie möglich anrichten. Allerdings gibt es mehrere Probleme in der Architektur dieser Maschinen. Da sie mit getaggten Daten gefüttert werden, ist eine große Menge an manuell getaggten Inhalten erforderlich. Da es diese zum Teil noch nicht gibt, nutzen Tech-Unternehmen einmal mehr den globalen Süden, um dank präkarisierter Arbeitskräfte in kürzester Zeit möglichst viele aufbereitete Daten zu erhalten und damit die Maschine füttern zu können, so Marcus Roeper. Dieses asymmetrische Machtverhältnis zum globalen Süden erinnert an alte koloniale Strukturen. Aber das ist noch nicht alles. Damit die Maschine lernen kann, sensible Inhalte von nicht sensiblen Inhalten zu unterscheiden, muss sie trainiert werden. Das bedeutet, dass riesige Mengen an sensiblen Inhalten von echten Menschen gelabelt werden müssen. Die Folgen dieser Arbeit, die meist unter höchst zweifelhaften Arbeitsbedingungen und ohne ausreichende psychologische Begleitung stattfindet, sind noch immer schwer abzuschätzen. Dieses Ungleichgewicht im Machtverhältnis zwischen Nord und Süd bzw. zwischen Arm und Reich nutzt eine durch koloniale Kontinuität und neoliberale Infrastruktur erzeugte Abhängigkeit aus, und lässt die betroffenen Menschen mit heute noch nicht abschätzbaren und unvorstellbaren Folgen allein zurück. Wie bei vielen anderen Produkten und Dienstleistungen in unserer globalisierten Welt stellen die Verbraucher*innen im Globalen Norden die mangelnde Transparenz der Unternehmen jedoch nicht in Frage oder akzeptieren sie sogar. Dabei nehmen wir ausbeuterische und unfaire Arbeitsbedingungen, Kinderarbeit oder Umweltverschmutzung in Kauf.

Eine weitere Frage, die sich mit der Informationslast bzw. der Einspeisung dieser neuen Technologien stellt, ist für Marcus Roeper auch die Frage nach der Anerkennung von Diversität und dem Entgegenwirken von Vorurteilen, Diskriminierung und Beleidigungen. Neutralität im digitalen Raum ist für ihn schwer vorstellbar. Bei den ersten Schritten im Bereich der künstlichen Intelligenz führte das Fehlen einer gründlichen Diversitätsarbeit dazu, dass Menschen of Color mit Affen verwechselt oder Frauen in Einstellungskontexten nicht für einen Job ausgewählt wurden, weil die Unternehmensdaten der Maschine zeigten, dass bis dato mehrheitlich Männer für die gesuchte Position eingestellt wurden und sie daraus schloss, dass nur Männer für den Job besser geeignet sind. Dieser Mangel an Sensibilität zur Förderung einer Pseudoneutralität offenbart strukturelle Ungereimtheiten in der Architektur der künstlichen Intelligenz. Zum einen muss die Datenlast eine Komponente der Inklusion und der wiederherstellenden sozialen Gerechtigkeit haben, um rassistische, hetero-cis-sexistische, ableistische, antisemitische, antimuslimisch-rassistische, migrantistische, coloristische, usw. Verhaltensweisen nicht zu wiederholen. Zum anderen zeigt sie uns, dass die Geschichte der Menschheit vielfach auf koloniale, rassistische und patriarchale Weise dokumentiert und katalogisiert wurde und wird und dass wir, weil die Politiken des Sammelns und der Wissensproduktion historisch gesehen monumental gewalttätig, diskriminierend und stigmatisierend waren, unsere eigene Geschichte kritisch aufarbeiten müssen, um nicht immer wieder die gleichen Fehler zu wiederholen.

Ein weiteres Problem, mit dem wir uns zukünftig auseinandersetzen müssen, ist für Marcus Roeper die Frage, wie wir mit den Informationen und dem Wissen, das durch künstliche Intelligenz entsteht, umgehen werden. Er erklärt, dass eine der ersten Fragen, die wir uns bei der Suche nach Informationen stellen, jene ist, welcher Quelle wir vertrauen können. Früher haben wir uns eine uns vertrauenswürdig erscheinende Zeitung angeschaut, heute vielleicht eine entsprechende digitale Zeitung oder ein Portal, von dem wir glauben, dass es mit intellektueller Redlichkeit gepflegt wird. Aber wenn uns etwas verdächtig oder unglaubwürdig erscheint, sind wir natürlich misstrauisch. In dem Kontext spielen Bilder eine wichtige Rolle. Früher glaubten wir, dass fotografische Bilder ein zuverlässiges Zeugnis dafür waren, dass etwas passiert ist und dass der*die Fotograf*in dabei war. Mit der Manipulation des fotografischen Bildes und dem heutigen systematischen Einsatz von Photoshop etc. haben wir begonnen, der Authentizität von Bildern zu misstrauen. Dennoch bedeutete ein fotografisches Bild auf die eine oder andere Weise mehr oder weniger genau, dass etwas tatsächlich passiert ist. Das Problem besteht nun darin, dass künstliche Intelligenz mittlerweile in der Lage ist, auf der Grundlage von Bildern und Texten, die real existieren, neue Texte und Bilder über Dinge zu erstellen, die noch nie existiert haben. Marcus Roeper gibt das Beispiel einer imaginären Explosion mitten in einer Stadt mit Rauch, Zerstörung und Hunderten von Schwerverletzten. Die Technik ist inzwischen so weit fortgeschritten, dass die künstliche Intelligenz nicht nur ein Bild einer Explosion in einer Stadt mit Verletzten erzeugen kann, sondern auch in der Lage ist, tausende von Bildern und sogar Videos über diese imaginäre Explosion zu kreieren. Innerhalb weniger Stunden könnte eine KI einen oder gar mehrere Blogs mit Artikeln über die fiktive Explosion erstellen, die in verschiedenen Stilen verfasst und mit unterschiedlichen selbst erstellten Bildern illustriert sind. Wie kann man diese scheinbar echten Nachrichten von einer Fiktion unterscheiden? Und wenn wir noch einen Schritt weiter denken: was wäre, wenn wir das Internet mit irrealen Geschichten, imaginären Bildern und Fiktionen überfluten, die in der Realität nicht existieren? Was ist, wenn diese Informationen von den seriösen Medien aufgegriffen werden? Google und andere Suchmaschinen könnten irgendwann nicht mehr in der Lage sein, sie zu ignorieren und diese Beispiele als mögliche Antworten anzuzeigen. Was in der Vergangenheit ein kleiner Blog war, der mit der Kraft einer einzigen Person und vielen Stunden Photoshop und redaktioneller Arbeit erstellt wurde, kann nun in wenigen Minuten „Realität" werden. Wir mögen denken, dass wir nicht von künstlicher Intelligenz beeinflusst werden können, aber es ist hochwahrscheinlich, dass wir bereits beeinflusst werden. Es fällt uns schon jetzt schwer, zu entscheiden, wem wir vertrauen und wie wir die Wahrheit erkennen können. Künstliche Intelligenz ist nicht darauf trainiert, uns zu sagen, was wahr ist, sondern

Inhalte so kohärent wie möglich zu schaffen und zu wiederholen. So wird es immer schwieriger, korrekte und echte Informationsquellen zu finden.

Um zu zeigen, wie einfach es ist, eine Maschine mit störenden Inhalten zu füttern, arbeitet Marcus Roeper derzeit an einem Modell, das aus Hassbotschaften lernt und personalisierten Hass für einzelne Nutzer*innen erstellen kann. Um zu zeigen, wie unkompliziert es ist, Hass zu automatisieren, und welche Gefahren wirklich dahinterstecken, nutzt er ein praktisches Beispiel, um das Bewusstsein für die Gefahren verbaler Gewalt zu schärfen. Er ist überzeugt, dass er nicht der Einzige ist, der an ähnlichen Modellen arbeitet, und hofft zugleich, dass die anderen Modelle ebenfalls zu pädagogischen Zwecken entwickelt werden, da, wie wir bereits wissen, die Verbreitung von Hass und der damit verbundene Anstieg an Nutzer*innen-Aktivität für Tech-Konzerne einen erhöhten Profit bedeuten kann.

Fake und Misleading
Über wie die Wahrheit, sofern es eine gibt

Stellen wir uns vor, im Internet eine von einem Menschen erstellt Website zu finden, die die absurde Behauptung aufstellt, dass ein Schwangerschaftsabbruch Krebs verursacht. Auf dieser Website gibt es Artikel, erklärende Bilder, Erfahrungsberichte von Menschen, die diese Behauptung bestätigen, usw. Nun stellen wir uns vor, dass eine künstliche Intelligenz diese Seite findet und diese Informationen reproduziert - und zwar tausendfach. Wenn wir an einen Punkt gelangen, an dem das Internet zunehmend mit Desinformationen, wissenschaftlich nicht belegten Fakten, Spekulationen und verschwörerischen Fantasien gefüllt wird, wird es sowohl für Suchmaschinen als auch für Menschen schwieriger werden, zwischen wahr und falsch zu unterscheiden. Nicht nur ChatGPT, sondern auch andere vergleichbare Modelle, die auf im Internet gesammelte Informationen trainiert sind, könnten beginnen, Unwahrheiten zu reproduzieren. Höchstwahrscheinlich tun sie das sogar bereits in diesem Moment. Aber das Problem wird noch komplexer. Die neuen Technologien der künstlichen Intelligenz unterscheiden nicht zwischen von Menschen erstellten Inhalten und maschinell erstellten Inhalten. Für KI ist alles Inhalt. Aus diesem Grund lernt sie auch von den Erfindungen, anderer KI's . Wenn die Maschine so eine falsche Realität schafft, nennt man das eine Halluzination. Von Menschen geschaffene Fakes oder maschinell erzeugte Halluzinationen zu stoppen, um Desinformation entgegenwirken, ist eine enorme Aufräumarbeit auf unvorstellbarem Niveau, die manuell von echten Menschen geleistet werden muss. DeepFakes, Fakenews und KI-Halluzinationen machen uns misstrauischer gegenüber unserer Wahrnehmung und schaffen eine Welt, in der wir immer weniger glauben können, was wir sehen, hören oder lesen. Wir leben in einer Welt, in der Wahrheit selbst fragwürdig ist, je nachdem, aus welchem Blickwinkel man sie betrachtet. Wenn wir dann noch die Komplexität der Unterscheidung zwischen Wahrheit und Lüge addieren, werden wir dringend neue Formen von Allianzen schaffen müssen, auf die wir uns verlassen können.

Zukunftsvorstellungen und Handlungs(un)möglichkeiten
Der Umgang mit einer Kraft, die wir noch nicht verstehen

Wir beginnen zu verstehen, dass wir es bei Hass im Netzt und künstlicher Intelligenz mit einer Kraft zu tun haben, die wir noch nicht verstehen und die wir auf jeden Fall nicht nur untersuchen, sondern auch lernen müssen zu kontrollieren. Dazu müssen wir ein angemessenes Vokabular entwickeln, das die Tiefe und Multidimensionalität der damit verbundenen Herausforderungen begreift, sowie eine Reihe von Instrumenten und Methoden, die es uns ermöglichen, von verschiedenen Ausgangspunkten aus die negativen Auswirkungen, die Angriffe auf die menschliche Sensibilität und Integrität zu kontrollieren und zu regulieren. Es sollten diversitätsorientierte Maßnahmen ergriffen werden, die intolerant gegenüber Hass sind und ein tiefes Verständnis für die unzähligen Arten der Sichtbarmachung von Diskriminierung haben, um die Architektur zu überdenken, die diese Technologien im digitalen Raum stützt. Zusätzlich zu einer Reihe von Maßnahmen, die die menschliche Vielfalt stärken und repräsentieren und Diskriminierung und Rassismus entgegenwirken, muss der künstlichen Intelligenz beigebracht werden, bewusst vorrangig von Menschen und nicht von anderen Maschinen zu lernen, denn es ist davon auszugehen, dass von Maschinen geschaffene Inhalte von Menschengeschaffene Inhalte quantitativ in kürzester Zeit übertreffen können. Dieses Ungleichgewicht in der Wissensproduktion entwirft ein dystopisches Szenario, in dem Wissen möglicherweise nur eine komplexe Ansammlung von Halluzinationen wird. Um dieser Gefahr entgegenzuwirken, müssen strenge Vorschriften auf der Grundlage der moralischen und ethischen Grenzen für das Training von Maschinen geschaffen werden, und es wird definitiv ein enormes Reinforcement learning from human feedback (Bestärkendes Lernen aus menschlichem Feedback) erforderlich sein, um eine stärkere Nutzung der künstlichen Intelligenz mit so wenig Kollateralschäden wie möglich zu ermöglichen. Dies wird natürlich die Schaffung gesunder Arbeitsplätze mit ausgebildeten Menschen erfordern, die zwischen der gigantischen Masse an Informationen und digitalen Inhalten, die wir mit in die Zukunft nehmen wollen, unterscheiden können. Es müssen zudem Lösungen gefunden werden, um den ganzen „digitalen Dreck" zu entsorgen, den die ungeregelten Tests in dieser Zeit des Herumexperimentierens hinterlassen. In jeder dieser Plattformen sowie in den multinationalen Konzernen, die hinter den KI's stehen, muss es starke und diversitätssensible Ethik-Teams geben, die die Menschen vertreten und über die tatsächlichen Auswirkungen beraten, wenn den Mechanismen der Gewalt, des Missbrauchs und der Unterdrückung innerhalb einer Plattform nicht entgegengewirkt wird. Auf einer anderen Handlungsebene sollten international staatliche Kommissionen zur Regulierung und Bestrafung von Hass und Desinformation geschaffen werden. Wie in anderen öffentlichen Räumen, wo es bereits Regeln des Zusammenlebens und Strafen für Menschen gibt, die die Bedingungen des Miteinanders nicht respektieren, sollten auch für den digitalen Raum Regeln und Verträge für Ordnung und Miteinander geschaffen werden. Früher oder später wird der digitale Raum Regeln für Koexistenz

brauchen, eine Garantie für Respekt. Gleichzeitig wird viel Aufklärungsarbeit im Bereich der digitalen Kompetenz geleistet werden müssen. Es muss Modelle wie das von Marcus Roeper geben, die zeigen, wie tief verletzend Hass und Angriffe im Netz sein können und wie gefährlich Desinformation werden kann. Dazu bedarf es einer intensiven Arbeit von Digitalexpert*innen und Pädagog*innen, die uns auf die kommende Welt vorbereiten. Auf der Ebene der Grund- und Sekundarschulen sollte Digital Litteracy eingeführt werden, in die Kinder und Jugendliche mit den tiefgreifenden sozialen und politischen Auswirkungen der digitalen Nutzung und Navigation in Berührung kommen. Sie sollten lernen, Fake News zu erkennen und Hass abzulehnen. Dafür sollte es einen wichtigen Platz im Klassenzimmer geben, um mit digitalen Werkzeugen wie Tablets, Smartphones und Computern, aber auch mit dem Internet, sozialen Netzwerken, Nachrichtenkanälen und Blogs in Kontakt zu kommen. Neben dem Erlernen technischer Fertigkeiten sollten Kinder und Jugendliche hierbei auch Regeln des Miteinanderlebens lernen, beispielsweise, wie man sich online verhalten kann, was wirklich verboten ist, wofür man angezeigt werden kann, was kritisch oder verletzend ist. Dies sind viele Aspekte, die bei den Kindern ankommen müssen, und Lehrer*innen (aber vor allem auch die darüber liegenden Strukturen, Behörden und Ministerien) müssen diesen Herausforderungen gerecht werden. Aber Bildung endet nicht bei den jüngeren Generationen, auch die Älteren werden auf globaler Ebene lernen müssen, mit diesen Technologien zu leben und den digitalen Analphabetismus zu überwinden. In einer eher utopischen Vorstellung könnte es eine obligatorische Grundausbildung im Umgang mit den digitalen Tools und Medien geben, um sich online so rücksichtsvoll zu verhalten, wie wir uns auch in anderen öffentlichen Bereichen verhalten sollten. Bislang sind alle Regulierungsversuche gescheitert. Wir müssen uns jedoch der Herausforderung stellen, sichere Räume zu schaffen, in denen sich die Menschen wohl fühlen, anonym bleiben dürfen und so die Technologie auf humane Weise nutzen können.

[1] Noble, Safiya Umoja (2018) Algorithms of Oppression: How Search Engines Reinforce Racism. New York: New York University Press.

Die Bedeutung von Hate Speech und der Umgang damit in der kritischen Bildungsarbeit

Ein Interview mit Žaklina Mamutović von Bildungsteam Berlin Brandenburg

Bahar Oghalai: Liebe Žaklina, kannst du das Bildungsteam Berlin-Brandenburg e.V. und eure Arbeit vorstellen?

Žaklina Mamutović: Das Bildungsteam feiert dieses Jahr das 25 jährige Jubiläum. Im Moment sind wir neun Teammitglieder. Wir versuchen möglichst viele Perspektiven in unsere Bildungsarbeit einzubeziehen.

Der Ursprung des Bildungsteams ist auch tatsächlich ein Politischer: Es ging um eine Unzufriedenheit darüber, wie Themen in der politischen Bildungsarbeit gesetzt werden und auch wie die Bezahlung in diesen Kontexten ist. Das Bildungsteam ist nach der Gründung sehr viel in Brandenburg aktiv gewesen und hat dort viel zum Thema Rechtsextremismus gearbeitet. In der Träger*innenschaft von Bildungsteam gab es auch das Projekt Bildungsbausteine gegen Antisemitismus. Und da ist auch ein Buch mit Methoden vor über 20 Jahren entstanden. Sie haben irgendwann festgestellt, dass sie, was Ost-West angeht, recht heterogen sind, auch was die Gender Positionierung angeht, aber dass eben alle Menschen, die dort arbeiten, weiß positioniert sind. Und das obwohl einige Projekte an Schulen in Kreuzberg und Neukölln angesiedelt waren und die Schüler*innenschaft eine ganz andere war/ist als was das Team damals repräsentierte. Das war dann eben eine Erkenntnis und dann haben wir nochmal sieben, acht BIPoC* Kolleg*innen angefragt und das vor siebzehn, achtzehn Jahren. Das hat dazu geführt, dass wir sowohl queere als auch heteronormative Positionierungen haben, dass wir BIPoC* Positionierungen und weiße Positionierungen haben. So waren wir als Team auch zusammengesetzt und das ist sozusagen auch der Punkt, an dem so ein Umdenken und eine Umstrukturierung in Bezug auf Herkunft im Bildungsteam stattgefunden hat. Dadurch werden auch Themen tatsächlich gerade in Bezug auf rassistische Diskriminierung und auch postkoloniale Perspektiven anders gesetzt. Da es eben nicht nur eine weiße Perspektive auf die Themen gibt, also die auch durchaus kritisch sein kann, aber jetzt bringt das Bildungsteam diese Perspektiven zusammen. In den letzten Jahren haben wir verstärkt an Schulen gearbeitet. Wir haben schon immer an Schulen gearbeitet. In den letzten Jahren wurde es aber immer wichtiger, dass wir Bildungsangebote für Jugendliche und junge Erwachsene machen, die eher weniger Zugang zur politischen Bildungsarbeit haben oder eben nicht in aktivistischen Kontexten unterwegs sind. Also sind wir immer weiter weg von Einzelseminaren und haben verstärkt Begleitungsformate zum Beispiel im Beruf gemacht. Dazu

haben wir dann eben auch Methoden entwickelt und haben den Schwerpunkt antimuslimischer Rassismus. Diese Formate werden intersektional konzipiert und da arbeiten wir an den Grundschulen mit den Lehrkräften, dem pädagogischen Personal, den Eltern usw. zusammen. Wir arbeiten auch mit Jugendämtern zusammen und begleiten die Öffnung der Jugendämter mit einem kritischen Blick.

BO: Sehr spannend. Also ihr macht kritische Bildungsarbeit und wir wissen, dass Institutionen, die kritische Bildungsarbeit machen, antirassistische Arbeit leisten, immer wieder eben auch von Hassrede oder von Hassinhalten betroffen sind. Wie sieht es da bei euch aus?

ŽM: Vielleicht nicht Hassrede, aber wir stoßen natürlich immer wieder auf diskriminierende Sätze und Aussagen im Rahmen unserer Bildungsarbeit.

Außerdem haben wir tatsächlich versucht, das Thema in dem Projekt zu antimuslimischem Rassismus zu bearbeite. Da haben wir auch eine Methode entwickelt, wo wir mit dem Song „Hey Mr. Sarrazin" von Kamyar und Dzeko arbeiten, wo es tatsächlich um die ganzen rassistischen Aussagen und die Hassrede von Sarrazin geht und ja quasi auch ein Song gegen Hate Speech ist. Und das haben wir zu einer Methode in unseren Bildungsangeboten verarbeitet.

BO: Und was macht diese Methode aus?

ŽM: Auf der einen Seite geht es um das Aufdecken von rassistischen Aussagen. Also, dass wir uns Passagen aus Sarrazins Buch „Deutschland schafft sich ab" anschauen und da herausarbeiten, was daran rassistisch ist. Wobei wir da schauen, je nach der Zusammensetzung von beispielsweise Klassen, ob das überhaupt notwendig ist. Es gibt natürlich Unterschiede in den Perspektiven von weiß positionierten Schüler*innen und jenen mit Rassismuserfahrung. Da sollte es einfach unterschiedliche Herangehensweisen geben. Denn wenn wir mit Schüler*innen arbeiten, die selbst von Rassismus betroffen sind, müssen sie sich diese verletzenden Inhalte natürlich nicht nochmal anhören. Also es geht um beides: auf der einen Seite um Empowerment bspw. mit dem Song von Kamyar und Dzeko, aber auch um Aufklärung und Herausarbeitung rassistischer Inhalte.

BO: Okay, und die diskriminierenden Aussagen, kommen diese auf Veranstaltungen vor?

ŽM: Die kommen in den Seminaren vor. Also das ist gar nicht mal so anonym. Das macht es ja auch nochmal so spannend. Viele denken Hate Speech hätte so einen starken Anonymitätscharakter. Aber uns stehen oft auch reale Personen gegenüber und geben hasserfüllte Kommentare von sich.

BO: Wie geht ihr damit um, wenn ihr Seminare gebt?

ŽM: Also ich bin ja der Meinung, dass Bildungsarbeit da aufhört. Es ist dann nicht mehr meine Aufgabe aufzuklären, wenn Menschen mich beleidigen. Da bin ich mir ganz bewusst, dass diese Menschen verletzen wollen. Sie wissen, was sie tun und müssen dafür auch Verantwortung tragen. Da versuche ich das auch nicht über den Kontext der Arbeit zu rechtfertigen und irgendwie meinen Bildungsauftrag in den Vordergrund zu stellen. Und da stellt sich natürlich auch die Frage, wie wir Hate Speech definieren, was wir als Hate Speech bezeichnen. Für mich gehört das schon auch zu Hate Speech, wenn insbesondere Kinder aus bestimmten Kontexten, sowie ihre Eltern Beleidigungen und rassistischen Aussagen seitens der Lehrkräfte ausgesetzt sind. Auch das ist Hate Speech. Und stell dir mal vor, wenn Menschen dir gegenüberstehen und dann wirklich ihre Tiraden loslassen, dann müssen sie sich so im Recht und sicher fühlen, dass ihnen nichts passiert. Und das ist besonders schlimm. Da sehen wir uns dann eher die Strukturen an, in Form von Beschwerden oder in sehr drastischen Fällen wird auch mit einer Anzeige gedroht, oder mit der Meldung beim Schulamt. Denn es gibt leider immer mal wieder Lehrkräfte, die uns aber auch Schüler*innen und Eltern gegenüber diskriminierende Aussagen treffen.

Ich weiß, dass es oft auch Beschwerden von Eltern von Kindern über sehr beleidigende Vorfälle gerade Müttern von muslimisch gelesenen Kindern gibt . Ich würde in diesem Zusammenhang auch gerne im Schulamt selbst Seminare anbieten, weil ich glaube, dass auch dort in diesen Strukturen noch sehr viel Sensibilisierungsbedarf besteht.

BO: Was bedeuten digitale Entwicklungen für euch? Im Allgemeinen findet ja mittlerweile sehr viel Bildungsarbeit einfach auch im Netz statt. Was bedeutet das für eure Arbeit?

ŽM: Ich merke es oft bei Veranstaltungen. Wir bieten oft digitale oder eben auch hybride Veranstaltungen an. Wir haben in dem Rahmen ein Schriftstück verfasst, das beinhaltet, was überhaupt nicht geduldet wird und wo Leute sofort ausgeschaltet werden, wenn sie diskriminierende Aussagen treffen. Gerade in der Zeit der Corona Pandemie, war ich selber bei der Organisation und Moderation von Veranstaltungen dabei, wo Störer*innen vortragenden Gäst*innen gegenüber sehr diskriminierende, hasserfüllte Aussagen getroffen haben. Das ist, das gebe ich zu, etwas, das mir wirklich Angst macht. Also auch die Sorge, wie wir so etwas dann in solchen Momenten unterbrechen können. Und aufgrund dieser Angst läuft alles per Anmeldung. Klar melden sich Menschen trotzdem auch nicht immer mit Klarnamen an. Und die Restsorge bleibt immer, dass du nicht weißt, wer wirklich hinter der Kamera steckt und welche Absichten hat. Ich versuche, die Kontrolle zu behalten und alle Funktionen zu nutzen, die den Vortrag und die Diskussion regeln und solche Zwischenfälle vermeiden. Wenn beispielsweise jemand eine Frage hat, dann geht sie nur an eine Person, die die Fragen verwaltet, damit keine Kommentare einfach so in den Chat geschrieben werden können. Das sind Maßnahmen, die uns digital zur Verfügung stehen. Wobei es natürlich auch immer wieder auf Präsenzveranstaltungen zu verletzenden und diskriminierenden Kommentaren kommen kann. Es können zwei Störer*innen eine ganze Veranstaltung sprengen. Deshalb müssen wir auch schon im Vorfeld darüber nachdenken, welche Riegel wir vorschieben können. Was wollen wir zulassen und was nicht, ohne dass uns das Ganze entgleitet. Dabei bietet der digitale Raum auch unglaubliche Möglichkeiten, eben Möglichkeiten der Reichweite aber auch Risiken, mit denen wir umgehen müssen.

Hass im Netz kommt aus allen Richtungen

Christina Hübers

Wie lernen Jugendliche, sich Hilfe zu holen, wenn sie Hass – oder Hatespeech – im Netz entdecken? Oder dem sogar selbst ausgesetzt sind? Ein Gespräch mit Christina Hübers vom Verein „ichbinhier“, der Workshops für digitale Zivilcourage anbietet.

Was genau ist Hatespeech?

Hatespeech – im Deutschen: Hassrede – ist nicht einheitlich definiert. Wir definieren sie als aggressive oder allgemein abwertende Aussagen gegenüber Personen, die bestimmten Gruppen zugeordnet werden. Besonders betroffen sind Menschengruppen, die aufgrund ihrer Hautfarbe, Herkunft, Religion, sozialen Situation oder ihres Geschlechts marginalisiert und diskriminiert werden. Gehetzt wird aber auch gegen Politiker*innen, Influencer*innen und Journalist*innen. Frauen sind dabei besonders oft Zielscheibe. Hassrede kommt häufig ‚von rechts‘, oft aus ‚Trollfabriken‘, und ist regelrecht organisiert. Letztlich kann sie aber aus allen Richtungen kommen, oder auch von Einzelpersonen.

Und was hat Hatespeech mit Demokratie zu tun?

Social Media ist ein öffentlicher Raum – also ein Raum, in dem Demokratie gelebt und gestaltet wird. Hass und Hetze bewirken, dass sich Betroffene aus diesem Raum und damit aus dem öffentlichen Diskurs zurückziehen und ihre Meinung dadurch öffentlich nicht mehr sichtbar ist. Das Fatale ist, dass die Wahrnehmung, was die Mehrheitsmeinung zu sein scheint, verzerrt wird und die Diversität unserer pluralen Gesellschaft nicht abgebildet wird. Wir finden, dass Social Media als demokratisch gestaltbarer Raum für alle zugänglich bleiben muss!

Was macht Ihr Verein „ichbinhier“ genau?

Wir sind ein Verein, der sich gegen Hass und Hetze im Netz und für eine starke Demokratie im Digitalen einsetzt. Seit 2016 haben sich über 40.000 Menschen im Rahmen der Online-Aktionsgruppe #ichbinhier auf Facebook ehrenamtlich mit Counterspeech gegen Hasskommentare und für Betroffene eingesetzt. Wir unterstützen die Arbeit dieser Gruppe. Aus der jahrelangen Praxis im Umgang mit Hass im Netz haben wir Bildungsformate entwickelt, um Menschen für den Umgang mit Hass im Netz zu sensibilisieren und ihnen zu zeigen, wie man dagegen vorgehen kann. Unsere Trainings- und Bildungsarbeit richtet sich an verschiedene Zielgruppen. Zum Beispiel an Kommunalpolitiker*innen, die oft Ziel von Hasskampagnen sind, oder auch an Schüler*innen. In unseren Schulworkshops, den ‚Bootcamps für digitale Zivilcourage‘, zeigen wir, wie man konkret gegen Hassnachrichten vorgehen kann. Und wie man ‚digitale Zivilcourage‘ lebt. Es geht darum, einen Werkzeugkasten zur Verfügung zu stellen, der alles enthält, was man in einer solchen Situation benötigt: zum Beispiel Informationen darüber, wo und wie man Hatespeech meldet. Oder Handlungsanweisungen, wie man auf Hass und Hetze reagiert, wenn man eine solche Attacke miterlebt.

Wie läuft so ein Bootcamp ab?

Zu Beginn des Workshops reden wir allgemein über Hass und überlegen gemeinsam, was eine Meinung ist und wann die Grenze der Meinungsfreiheit erreicht ist. Der Kern unseres Bootcamp-Angebots ist die Simulation einer eskalierenden Kommunikation. Dafür weisen wir den Teilnehmenden verschiedene Online-Rollen zu, die man auch sonst in Online-Diskussionen findet: Es gibt auf der einen Seite User*innen, die sich eine Meinung bilden wollen, daneben User*innen, die sich austauschen und mit anderen offen diskutieren wollen. Auf der anderen Seite gibt es diejenigen, die nur an destruktiver Kommunikation interessiert sind. Diese Rollen treffen in unserer Simulation

aufeinander. Dazu stellen wir einen geschützten Raum auf einer Plattform zur Verfügung, die einem sozialen Netzwerk nachgebildet ist. Dann wird in der Simulation über ein ‚Trigger'-Thema gesprochen, das sich die Klasse zuvor ausgesucht hat und Diskussionspotenzial bietet. Alle gehen ihren Rollenbeschreibungen entsprechend vor – und schon bald eskaliert die Diskussion, während andere versuchen, dies zu verhindern. Anschließend wird dieses Online-Rollenspiel gemeinsam reflektiert: Wir fragen die Jugendlichen beispielsweise, wie sie ihre Rolle erlebt und welche Handlungsstrategien sie angewendet haben.

Zu welchen Aha-Erlebnissen führt das?

Den Schüler*innen wird oft zum ersten Mal bewusst, wie einfach es ist, im Netz destruktiv zu sein. Gleichzeitig verstehen sie, wie anspruchsvoll es sein kann, einen konstruktiven Online-Beitrag zu verfassen. Sie erleben zudem, dass man Hass mit Hilfe von eingeübten Strategien begrenzen kann. Die Person, die Troll war, muss nicht viel nachdenken. Sie kann Hass wie am Fließband produzieren. Durch die Reflexion dieser Rolle versteht man, dass sich der ausgeschüttete Hass nicht persönlich an eine Person richtet, sondern ins Netz gekippt wird, um damit Aufmerksamkeit zu bekommen, indem möglichst viele Menschen darauf reagieren.

Was lernen sie noch?

Sie lernen, unter welchen Bedingungen und mit welchen Mitteln man eine entgleisende Diskussion retten kann – zum Beispiel, indem man sich Verbündete sucht. Um sich selbst zu schützen, kann es auch zielführend sein, aus einer Endlosdiskussion einfach auszusteigen und sich zurückzunehmen. Wir zeigen zudem, wo man Hatespeech melden kann, weil sie Grenzen überschreitet und deshalb verboten ist. Und wie man Plattformbetreiber*innen auffordert, eine Diskussion zu moderieren.

Ab welchem Alter werden die Bootcamps angeboten?

Unsere Erfahrung zeigt, je älter die Schüler*innen sind, umso leichter fällt es ihnen, die Rollenbeschreibungen für den Online-Konflikt anzuwenden. Bislang haben wir die Simulation deshalb insbesondere ab der neunten Klasse angeboten. Wir sind gerade dabei, sie in angepasster Form auch für die Jahrgänge sieben und acht zugänglich zu machen.

Zu guter Letzt: Wie grenzt man Cybermobbing von Hatespeech ab?

Bei Cybermobbing richtet sich der Hass gezielt gegen eine Einzelperson, die den*die Aggressor*in aus dem privaten Umfeld zumeist auch persönlich kennt. – Hass im Netz hingegen richtet sich wahllos gegen Menschengruppen oder Personen des öffentlichen Lebens.

Wenn das Bootcamp dann vorbei ist – sind alle Schüler*innen Netz-Demokrat*innen?

Unser Einsatz kann nur ein Anfang sein, um Schüler*innen für das Thema zu sensibilisieren. Es wäre wichtig, dass Lehrkräfte das als Startschuss verstehen und darauf aufbauen! Aber grundsätzlich sind wir davon überzeugt, dass man digitale Zivilcourage erlernen kann.

Wie kann man ein Bootcamp buchen?

Bootcamp-Trainer*innen arbeiten zurzeit in Hamburg, Schleswig-Holstein, Bremen und dem nördlichen Niedersachsen, Berlin und Brandenburg. Schulen können uns direkt anfragen. Wer Interesse hat, meldet sich gerne jederzeit. Falls keine Mittel vorhanden sind, schauen wir nach einer finanziellen Lösung.

Glossar

Hatespeech: Unter Hatespeech versteht man die Äußerung von Hass durch gezielte Herabwürdigung, Beleidigung und Bedrohung einzelner Personen oder Personengruppen.

Meinungsfreiheit: Nach Artikel 5 des deutschen Grundgesetzes hat jeder das Recht, seine Meinung frei zu äußern und zu verbreiten und sich aus allgemein zugänglichen Quellen ungehindert zu informieren.

Cybermobbing bezeichnet jegliche Form des Herabwürdigens anderer im Internet oder über Smartphones.

Aus scout. Das Magazin für Medienerziehung https://www.scout-magazin.de/bildung-und-wissen/artikel/hass-im-netz-kommt-aus-allen-richtungen.html (20.07.2023)

Un-Learning Gegenrede in den sozialen Medien

Monika Hübscher

Dass sich unsere Gesellschaft über soziale Medien vernetzen kann, geht nicht ohne erhebliche Nebenwirkungen einher, und die schwerwiegenden gesellschaftlichen Auswirkungen von Hass auf den algorithmisch organisierten Plattformen sind mittlerweile weithin bekannt. Dies hat die Zivilgesellschaft dazu veranlasst, nach angemessenen Lösungen für die Bewältigung dieses Problems zu suchen. Aktuell gibt es zwei bedeutsame Ansätze, mit denen Nutzer*innen Hass in sozialen Medien entgegentreten können: Einerseits besteht die Möglichkeit, einen hasserfüllten Beitrag bei der Plattform zu melden, während andererseits die Option besteht, sich durch einen Kommentar unter einem hasserfüllten Beitrag zu positionieren.

Das Melden von Kommentaren auf den Plattformen ist komplex, und die niedrige Erfolgsrate lässt Nutzer:innen schnell resignieren. Im Gegensatz zur Meldung und der potenziellen anschließenden Löschung von Hassrede wird durch die Gegenrede der Hass jedoch lediglich konterkariert. Doch das vor allem das in Deutschland weit verbreitete Mittel der Wahl - die Gegenrede – bereitet Probleme, da ein Kommentar unter einem hasserfüllten Post den Hass im Grunde weiterverbreitet. Die Engagement-Algorithmen der sozialen Medien bevorzugen Inhalte, die ein hohes Maß an Interaktion aufweisen, wie beispielsweise Likes, Kommentare und Shares. Wenn ein Beitrag viele Kommentare erhält, signalisiert dies den Algorithmen, dass der Beitrag kontrovers oder relevant ist, was dazu führen kann, dass er einer breiteren Zielgruppe angezeigt wird. Wird also unter einem hasserfüllten Beitrag kommentiert, bekommen besonders viele Nutzer:innen diesen zu sehen. Dies führt dazu, dass vor allem problematische Inhalte auf sozialen Medien eine hohe Nutzer:innenaktivität aufweisen und besonders präsent sind. Zum Beispiel könnten Nutzer:innen, die antisemitische Inhalte teilen und dafür Likes und Shares erhalten, sich in ihrer hasserfüllten Position bestätigt fühlen. Gleichermaßen könnten Nutzer:innen, die auf hasserfüllte Beiträge stoßen, dazu tendieren, diese als wahr anzusehen, wenn sie feststellen, dass solche Beiträge rege diskutiert werden. Durch Gegenrede unter hasserfüllten Posts kämpft man nicht gegen Hass, sondern gegen Algorithmen.

Zudem stellt die automatisierte Verbreitung von Hate Speech durch Bots und Künstliche Intelligenz (KI) die Wirksamkeit von Gegenrede auf der Ebene der menschlichen Interaktion zusätzlich infrage. Es muss auch kritisch hinterfragt werden, warum die Verantwortung für ein technologisch erzeugtes Problem bei den Nutzer:innen liegen sollte. Denn jede Form von Gegenrede wirkt im Sinne der Anbieter:innen von sozialen Medien, da sie Interaktion auf den Plattformen generiert und deshalb kommerzialisierbar und profitabel ist.

Social Media Literacy gegen Hass

Dass Hass in den sozialen Medien trotz Bildungsprogrammen zur Gegenrede in den letzten Jahren zu einem immer größeren Problem geworden ist, spricht dafür, die Wirkungsweisen von Algorithmen und Features wie (Dis-)Likes, Kommentare und Shares in die Bildungsbemühungen miteinzubeziehen. Was bisher eher weniger der Fall ist.

Um auf Nutzer:innenebene dazu beizutragen, Hass in den sozialen Medien wirksam zu reduzieren, ist das Verständnis der Technologie ebenso wie die Identifizierung und Dekonstruktion von hasserfüllten Inhalten unerlässlich. So müssen beispielsweise für die Arbeit gegen Antisemitismus in sozialen Medien Kompetenzen zur Dekonstruktion von antisemitischen Inhalten und Social Media Literacy ineinandergreifen. Der Workshop "Social Media Literacy gegen Antisemitismus" im Projekt "Antisemitismus und Jugend" der Universität Duisburg-

Essen hilft Nutzer:innen nicht nur dabei, Antisemitismus zu erkennen, sondern auch die Technologie sozialer Medien zu verstehen und somit Hass auf den Plattformen zu reduzieren.

Unter Social Media Literacy (SML) verstehen wir vom Projektteam „Antisemitismus und Jugend“ die Fähigkeit der Nutzer:innen, Inhalte in den sozialen Medien aus technischer, kognitiver und emotionaler Sicht kritisch zu bewerten. Aus technischer Sicht umfassen SML Themen wie die Rolle von persönlichen Daten, Algorithmen und gezielter Werbung bei der Verbreitung von Antisemitismus und deren Auswirkungen auf die Gesellschaft. Auf der kognitiven Ebene beinhaltet SML die Fähigkeit, trotz sozialer Validierung durch Likes und Kommentare und einer hohen Anzahl von Followern glaubwürdige Quellen in einer Social Media-Umgebung zu unterscheiden. Dazu gehört auch die Fähigkeit, Hassreden und Desinformationen im Zusammenhang mit dem Holocaust zu erkennen. Auf einer emotionalen Ebene umfasst SML die Fähigkeit, angemessen auf antisemitische Inhalte in sozialen Medien zu reagieren und zu interagieren.

Um die Widerständigkeit gegen Antisemitismus in die SML zu integrieren, lernen Nutzer:innen, gängige Erzählmuster und Stereotypen, die in antisemitischen Inhalten verwendet werden, zu erkennen und die Quellen und Motivationen hinter solchen Inhalten zu identifizieren. Dazu gehört auch die Erkenntnis, wie Antisemitismus und andere Formen der Diskriminierung, wie Rassismus, Misogynie und Hass gegen die LGBTQI+ Community, ineinander übergreifen.

Generell sollte SML eine Diskussion über eine verantwortungsvolle Nutzung sozialer Medien umfassen, wie zum Beispiel die Vermeidung des Teilens von und Kommentierens unter hasserfüllten Inhalten. Wenn Nutzer:innen besser in der Lage sind, Hass in sozialen Medien zu erkennen und zu verstehen, tragen sie zu einer sichereren und inklusiveren Social-Media-Erfahrung bei.

Die gute Form der Gegenrede: Verbundenheit, Solidarität und Dekonstruktion

Anstatt unter einem hasserfüllten Post zu kommentieren, können Nutzer:innen einen eigenen Beitrag verfassen und auf diese Weise Verbundenheit und Solidarität mit den von Hass Betroffenen zum Ausdruck bringen. Ein solcher Beitrag könnte beispielsweise lauten: "Ich stehe in Solidarität an der Seite der jüdischen Community und fordere ein Ende der antisemitischen Gewalt!". Der Post könnte dann mit dem offiziellen Profil der jüdischen Gemeinde verlinkt werden. Dies setzt nicht nur ein klares Zeichen gegen Hass, sondern kann auch dazu beitragen, dass die Solidaritätsbotschaft durch Kommentare unter dem Beitrag weite Verbreitung findet. Auf diese Weise wird Hass nicht weiterverbreitet, sondern eine positive Botschaft gegen den Hass gesetzt.

Minorisierte Gruppen wie die Roma-Community, die Trans-Community und Menschen, die mit Behinderungen leben, sind besonders von Hass in den sozialen Medien betroffen. Sich mit ihren Social-Media-Profilen zu verbinden und ihre Beiträge zu liken und zu teilen, stellt ebenfalls eine Form von Gegenrede dar, die zudem für Verbundenheit sorgt.

Bildungsangebote können Social Media Literacy gegen Hass anbieten, einschließlich der Dekonstruktion von hasserfüllten Inhalten. Die Plattformen der Bildungsangebote könnten anonymisierte Hassbeiträge als solche markieren (mit einem roten Verbotszeichen), diese dekonstruieren und sicher sowie kontrolliert veröffentlichen. Dadurch würde nicht nur ein Bildungseffekt für Social-Media-Nutzer:innen direkt auf den Plattformen entstehen, sondern gleichzeitig würden Hasspostings etwas entgegengesetzt und Solidarität mit betroffenen Gruppen gezeigt.

Weitere Literatur zum Thema

Antisemitism on Social Media
Monika Hübscher und Sabine von Mering (Hg.)
https://www.routledge.com/Antisemitism-on-Social-Media/Hubscher-Mering/p/book/9781032059693#

There is a lot of antisemitic hate speech on social media – and algorithms are partly to blame
Sabine von Mering und Monika Hübscher
https://theconversation.com/there-is-a-lot-of-antisemitic-hate-speech-on-social-media-and-algorithms-are-partly-to-blame-185668

Zum Schluss ein Plädoyer: neue Allianzen schmieden

María do Mar Castro Varela

Die Pandemie hat uns mit bedeutenden Fragen konfrontiert und gleichzeitig die Fragilität der Demokratie und ihrer Institutionen offengelegt. In Berlin versammelten sich zu Spitzenzeiten 20.000 Menschen, um gegen staatlich verordnete Maßnahmen zur Eindämmung der Pandemie zu protestieren und für ihre Freiheitsrechte einzutreten. Unter den vielfältigen Protestierenden fanden sich nicht zufällig viele rechte Stimmen und Positionen, darunter Reichsbürger:innen und Querdenker:innen. Demokratische Prinzipien wurden dabei vereinzelt mit dem „Nazistaat" verglichen, und die Maßnahmen gegen die Pandemie wurden mit der Verfolgung jüdischer Menschen oder den Stasimethoden der DDR gleichgesetzt. Antisemitismus und Rassismus erlebten einen neuen Höhepunkt, begleitet von florierenden Verschwörungstheorien. Ob es um „Die Weisen von Zion", den „Großen Austausch" oder rassistische Erzählungen von einer „gelben Gefahr" ging – sämtliche denkbaren und undenkbaren Verschwörungstheorien kursierten und kursieren weiterhin im Netz. Sie werden wild kombiniert mit oft absurden Begründungen, die entweder leugnen, dass das COVID-19-Virus existiert, oder behaupten, es wurde bewusst verbreitet, um bestimmte Gruppen an die Macht zu bringen.

Prägnant auf den Punkt gebracht: Wir befinden uns in einer Zeit, die von Antiaufklärung geprägt ist. Über soziale Medien verbreiten sich die absonderlichsten Nachrichten im Eiltempo, und je mehr sie verbreitet werden, desto glaubwürdiger scheinen sie zu werden. Das Denken wird schnell instabil und verfällt in Wahn, sobald größere Unsicherheit ins Leben tritt und viele Menschen kaum glaubhafte Ideen verbreiten. Es wurde deutlich, dass viele Menschen in Deutschland der Wissenschaft, der Politik, dem Staat und den Medien nicht vertrauen. Dies sind zwar grobe Analysen, aber es wurde nicht nur offensichtlich, dass Menschen in Europa, wenn sie mit einer großen Krise konfrontiert sind, schnell Verschwörungsideologien verfallen; es wurde auch sichtbar, dass sie dann eher bereit sind, Hass gegen geflüchtete Menschen, Jüd:innen oder rassifizierte Menschen zu verbreiten. Zu Beginn der Pandemie waren es beispielsweise zunächst anti-asiatische Hassreden, die sich dann jedoch rasch auf andere Gruppen ausweiteten. So wurden in den sozialen Medien türkischstämmige und arabische Menschen für die Ausbreitung der Pandemie verantwortlich gemacht.

Die Studie zu digitalem Hass während der COVID-19-Pandemie, die sich u.a. die Frage gestellt hat, warum und wie es zu einem solch' rapiden Anstieg von Hassreden kommen konnte, verdeutlichte, wie soziale Medien alte Verschwörungsnarrative nutzten, um Angst zu schüren und Empörung zu erzeugen. Die politische Kommunikation von Regierungen mit ihren Bürger:innen gestaltete sich oft problematisch, und die Social Media-Plattformen waren selten bereit, Kontrollmechanismen einzuführen. Hassreden und schillernde Verschwörungstheorien fungieren bekanntlich als effektive Clickbaits, weswegen sie ungern von den Plattformen beseitigt werden.

Insbesondere die Diskursanalysen, die im Rahmen des Projekts geführt wurden, deuten auf eine bedenkliche Fragilität der Demokratie hin. Die Ergebnisse der Analysen stimmen mit zahlreichen anderen Studien überein, die darauf hinweisen, dass der Rechtspopulismus im Internet während der Pandemie weiterhin an Bedeutung gewonnen hat. So lesen wir in einem Online-Beitrag von HateAid im April 2023:

> „Die Zeiten, in denen Rechtsextremismus sich vor allem auf der Straße abspielte und an Bomberjacken und Glatzen erkennbar war, sind längst vorbei. Mittlerweile geben sich Rechtsextremist*innen nicht mehr leicht zu erkennen. Und sie haben das Internet zum wichtigsten Ort für die Verbreitung ihrer Ideologien und Verschwörungsmythen gemacht.“[1]

Angesichts der raschen Veränderungen in den technischen Rahmenbedingungen ist es unvermeidlich, Diskursanalysen und die politische und kulturelle Vermittlungsarbeit den sich verändernden Gegebenheiten anzupassen. So richtete das Projekt eine besondere Aufmerksamkeit etwa auf neue Subjektformationen, wie zum Beispiel Prosumer:innen und Influencer:innen, sowie auf neue Dialogpartner:innen wie Bots und Algorithmen. Sie alle spielten eine außerordentliche Rolle bei der Verbreitung von Hassreden – und leider viel seltener bei der Eindämmung derselben.
Die Analysen zeichneten nach, wie alte Hassreden durch den Einsatz neuer Medienmittel zeitgemäß verbreitet wurden – sei es durch Memes, Tik-Tok-Videos oder Podcasts. Das desinformierende und gewalttätige Gezwitscher sorgte dabei für verstörendes Entertainment. Bereits 2002 führte Lisa Nakamura das Konzept der Cybertypen ein, das rassistische Online-Repräsentationen beschreibt. Nakamura analysiert, wie diese Darstellungen im Internet konstruiert werden, und hebt hervor, dass der Cyberspace keine neutrale Sphäre, sondern von kulturellen und sozialen Bedeutungen durchdrungen ist. Diese Beeinflussungen wirken sich darauf aus, wie rassifizierte Identitäten online verstanden und erlebt werden. Nakamuras Arbeit beleuchtet die Kontrolle dominanter Gruppen über die Darstellung rassifizierter Subjekte, was dazu führt, bestehende Machtstrukturen zu verstärken und bestimmte soziale Gruppen (etwa Migrant:innen) zu diskriminieren. Gleichzeitig blickt die Studie auf den Widerstand und die Gegenentwürfe von marginalisierten Gruppen, die diese dominierenden Darstellungen herausfordern und unterminieren.
Eine Zunahme und Dynamisierung staatsphobischer Diskurse ist deutlich auszumachen / zu erkennen. Staatsphobischen Diskursen wurde in kritischen Analysen bisher zu wenig Aufmerksamkeit geschenkt. Angesichts ihrer Dominanz während der COVID-19-Pandemie und der Beschleunigung von Desinformation, die mit dieser einhergeht, ist es dringend geboten, sie genauer zu untersuchen.
Der Begriff „Staatsphobie“, geprägt von Michel Foucault (2006), bezieht sich auf vereinfachende Erklärungen für die Funktionsweise staatlicher Macht. Im staatsphobischen Diskurs wird der Staat als „Feind“ oder „Dämon“ dargestellt, den es zu bekämpfen oder zu überwinden gilt. Nikita Dhawan (2020) argumentiert überzeugend, dass Staatsphobie problematisch ist, da die fortwährende Dämonisierung des Staates einen gefährlichen Riss im politischen Leben erzeugen kann, der potenziell Raum für die Verbreitung antidemokratischer und rechtsextremer Ideologien schafft. In einer Studie zu auf Instagram kursierenden Verschwörungstheorien stellen Tuters und Willaert (2022) beispielsweise fest, dass Staatsphobie verschiedene politische Meinungen und Ideologien zusammenführen kann. Sie bildet eine gemeinsame Grundlage, auf der Menschen mit unterschiedlichen ideologischen Hintergründen gemeinsam gegen den Staat auftreten, indem sie sich selbst als „kritisch“ repräsentieren (Tuters/Willaert 2022, S. 1233).
Im Frühjahr 2020, als die deutsche Politik mit der drängenden Frage konfrontiert wurde, wie sie mit der raschen Ausbreitung der Pandemie umgehen sollte, reagierte die Öffentlichkeit uneinheitlich auf Maßnahmen wie Abriegelungen, Grenzschließungen und die Verpflichtung zum Tragen von Masken. Insbesondere in Regionen mit stärker verbreiteten rechtsextremen politischen Einstellungen wurde eine generelle Skepsis gegenüber öffentlichen Gesundheitsmaßnahmen beobachtet. Die spärlichen und oft widersprüchlichen Informationen über Ursachen, Auswirkungen und Verbreitung des Virus wurden in populistischen Diskursen als Beweis für staatliche Inkompetenz oder Böswilligkeit interpretiert. Dies bildete einen günstigen Nährboden für die Produktion und Verbreitung von Fehlinformationen, Desinformationen, Verschwörungstheorien und Hassreden. Digitale Medien ermöglichten dabei eine besonders schnelle Verbreitung irreführender Informationen.
In seiner Studie über die Mobilisierung von Emotionen durch rechtsgerichtete Gruppen stellt Strick (2021) aufschlussreich fest, dass rechtsgerichtete Diskurse zwar immer noch oft als eine Form der extremen Rede angesehen werden, aber in Wirklichkeit bereits Teil des Mainstreams und der Alltagskommunikation sind – insbesondere in den sozialen Medien.

Darüber hinaus unterstützen die sozialen Medien die Verschärfung bestehender Wir/Sie-Unterscheidungen, insbesondere im Angesicht einer unsichtbaren Bedrohung – wie etwa eines Virus. Um ein Gefühl der Zusammengehörigkeit gegen einen unsichtbaren Feind zu schaffen, erklärten Politiker dem Virus den Krieg, wie in Frankreich oder verdrängten das Virus diskursiv über die Landesgrenzen hinaus, wie bei Donald Trumps falsche Benennung des „chinesischen Virus“. Solche diskursiven Strategien beschwören ein auf der nationalen Identität basierendes Gefühl der Einheit herauf, was unweigerlich Fragen aufwirft wie die, wer zur Nation gehört, wer Teil von „wir, das Volk“ ist und wer als Bedrohung angesehen wird.

Letztlich kann die Studie nur einen kleinen Einblick in das geben, *wie* sich politische Kommunikation mit den sozialen Medien verändert und dem *wer* von den differenten sozialen Medien profitiert. Klar wurde jedoch, dass neue Strategien des Umgangs mit rechtspopulistischen, rassistischen und antisemitischen Inhalten gefunden und die Vermittlungsformate politischer und kultureller Bildung darauf reagieren müssen. Es scheint uns unerlässlich, dass hier neue Allianzen zwischen kritischen KI-Expert:innen, Coder:innen, Influencer:innen und etwa rassismuskritischen kulturellen und politischen Vermittler:innen geschmiedet werden.

[1] Siehe https://hateaid.org/rechtsextremismus-im-internet/ (23.11.2023)

Literatur

Dhawan, Nikita (2020): “State as pharmakon”, in Davina Cooper/Nikita Dhawan /J. Newman (Hrsg.): Reimagining the State. Theoretical Challenges and Transformative Possibilities. London/New York: Routledge, S. 57–76.

Foucault, Michel (2006): Vorlesungen am Collège de France 1977-1978. Frankfurt a.M.: Suhrkamp.

Nakamura, Lisa (2002): Cybertypes. Race, Ethnicity, and Identity on the Internet. New York/London: Routledge.

Strick, Simon (2021): Rechte Gefühle. Affekte und Strategien des digitalen Faschismus. Bielefeld: transcript.

Tuters, Marc/Willaert, Tom (2022): “Deep state phobia: Narrative convergence in coronavirus conspiracism on Instagram”, in: Convergence: The International Journal of Research into New Media Technologies, 28(4), S. 1214–1218. https://doi.org/10.1177/13548565221118751.

Media

Video - 0:20:36
Digitaler Hass: Einführung
María do Mar Castro Varela und Helena Mihaljević
Deutsche Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96159

Symposium „Digitaler Hass“
im Haus der Kulturen der Welt

Tag 1 - 29.9.2022

1 - Digitaler Hass: Einführung

2 -The Right to Provoke? Free Speech, Hate Speech and the Politics of Censorship

3 - Rechte Diskurse on- und offline

4 - Digitaler Hass - Die Rolle sozialer Medien und digitaler Plattformen

Tag 2 - 30.9.2022

1 - „Driving on the Right“: Analyzing the Interdependence of Politics and the Media

2 - Gendered Disinformation and Computational Propaganda in Brazil: a Permanent Campaign

3 - Die Rolle der Algorithmen: von der Empfehlung bis zur Moderation von Inhalten

4 - Künstliche Intelligenz in der Kunst –Eine Strategie der politischen Bildung

5 - “Driving on the Right”: Analyzing the Interdependence of Politics and the Media

6 - Die Rolle der Algorithmen: von der Empfehlung bis zur Moderation von Inhalten

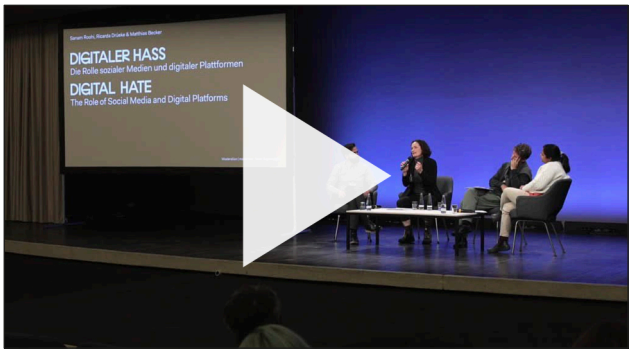
7 - Stark im Kampf gegen Hass im Netz - Die Rolle der Zivilgesellschaft

Tag 1 und 2

Lehrer*innenfortbildung unter der Leitung von ichbinhier e.V. (Auf die Fortbildungen folgten Schulprojekttage mit ichbinhier e.V. an vier Berliner Schulen).

Video - 0:57:39
The Right to Provoke? Free Speech, Hate Speech and the Politics of Censorship
Nikita Dhawan - Moderation María do Mar Castro Varela
Englische Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96160

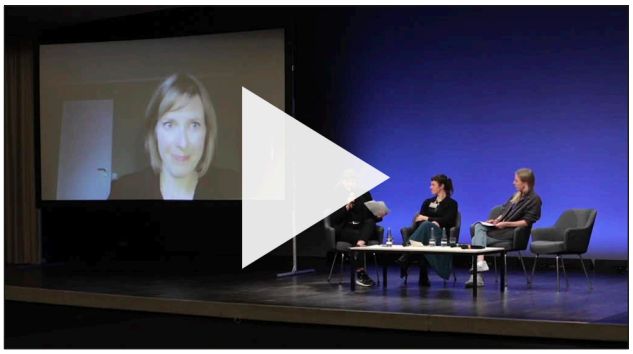
Video - 1:21:48
Rechte Diskurse on- und offline
Monika Hübscher, Kien Nghi Ha und Grischa Stanjek - Moderation Puneh Abdi
Deutsche Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96162



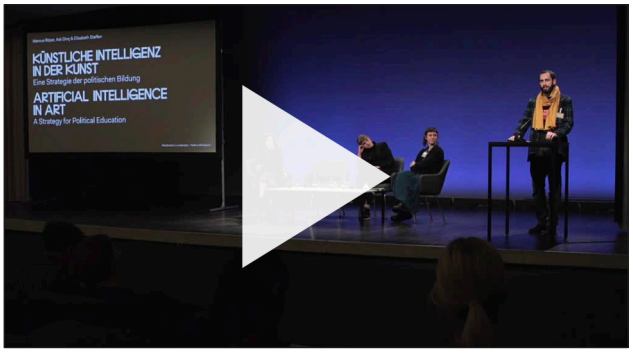
Video - 1:04:50
Digitaler Hass - Die Rolle sozialer Medien und digitaler Plattformen
Sanam Roohi, Ricarda Drüeke und Matthias J. Becker - Moderation Yener Bayramoğlu
Deutsche und Englische Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96191

Video - 0:42:08
“Driving on the Right”: Analyzing the Interdependence of Politics and the Media
Ruth Wodak - Moderation Helena Mihaljević
Deutsche Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96192

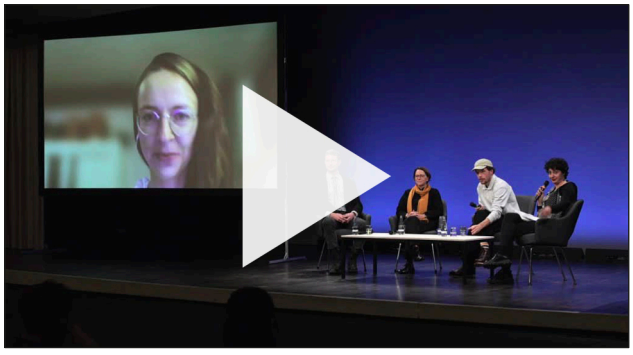
Video - 0:47:38
Gendered Disinformation and Computational Propaganda in Brazil: a Permanent Campaign
Rose Marie Santini - Moderation Helena Mihaljević
Englische Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96193



Video - 0:51:38
Die Rolle der Algorithmen: von der Empfehlung bis zur Moderation von Inhalten
Anne Kaun, Helena Mihaljević und Leah Nann - Moderation Milena Pustet
Deutsche und Englische Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96194



Video - 0:45:44
Künstliche Intelligenz in der Kunst – Eine Strategie der politischen Bildung
Aslı Dinç, Marcus Röper und Elisabeth Steffen - Moderation Helena Mihaljević
Deutsche Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96195



Video - 1:08:29
Stark im Kampf gegen Hass im Netz – Die Rolle der Zivilgesellschaft
María do Mar Castro Valera, Daniel Neugebauer, Chris Wagner, Stefanie Zacharias - Moderation Bahar Oghalai
Deutsche Originalversion
https://archiv.hkw.de/de/app/mediathek/video/96196

Bios

Autor*innen

María do Mar Castro Varela ist Diplom-Psychologin, Diplom-Pädagogin und promovierte Politikwissenschaftlerin. Sie ist Professorin für Allgemeine Pädagogik und Soziale Arbeit an der Alice Salomon Hochschule Berlin sowie Mitglied der Forschungsgruppe »Radiating Globality« und hat 2017 das bildungsLab* gegründet.

Helena Mihaljević ist Professorin für Data Science an der Hochschule für Technik und Wirtschaft in Berlin. Sie analysiert Daten und Technologien und wendet dabei Methoden der Data Science, des Maschinellen Lernens sowie des Natural Language Processing an. Sie forscht in inter- und transdisziplinären Projekten, beispielsweise zu algorithmischer Erkennung von Verschwörungstheorien und antisemitischer Hassrede. Vor ihrer Professur an der HTW war sie als Senior Data Scientist sowie im Bereich wissenschaftlicher Informationsinfrastruktur tätig. Ihre Promotion in Mathematik befasste sich mit der topologischen Dynamik ganzer transzendenter Funktionen.

Puneh Abdi ist Projektmitarbeiterin bei democ. im Projekt #TellMeMore. Zuvor hat die Soziologin (BA) als wissenschaftliche Mitarbeiterin im Projekt "Digitaler Hass" Online-Hassrede, Verschwörungsideologien und Antisemitismus im Kontext der Covid-19 Pandemie untersucht. Ihre Forschungsschwerpunkte sind (sub-)kulturelle Erscheinungsformen von Antisemitismus und Verschwörungsideologien. Gegenwärtig studiert sie Musikwissenschaft (MA) und forscht zu Antisemitismus im Gangstarap.

Christina Hübers ist Geschäftsführerin des ichbinhier e.V.. Seit September 2021 unterstützt sie die Arbeit des Vereins mit ihren umfangreichen Erfahrungen im Non-Profit Management, Communitybuilding und der Ehrenamtsarbeit

Monika Hübscher promoviert an der Universität Haifa in Israel und arbeitet außerdem als wissenschaftliche Mitarbeiterin für das Projekt „Antisemitismus und Jugend“ an der Universität Essen-Duisburg in Deutschland. Für ihre Forschung erhielt sie Stipendien des Deutschen Akademischen Austauschdienstes (DAAD) und der Fondation pour la Mémoire de la Shoah. Sie war Gastwissenschaftlerin an der Universität Bielefeld und der Universität Aarhus.

Bahar Oghalai ist Sozialwissenschaftlerin und forscht zu Politisierungsbiografien migrantischer Feminist*innen aus dem Iran. Sie forscht zu Intersektionen von Feminismus und Rassismuskritik. Sie ist außerdem wissenschaftliche Mitarbeiterin an der Alice-Salomon-Hochschule. Sie publiziert regelmäßig zu den Themen Feminismus und Migration mit einem besonderen Fokus auf die WANA-Region..

Verónica Orsi ist Kuratorin, Dozentin und Beraterin für kritische Diversität und Inklusion, feministische und hegemoniekritische Bildung und diversitätssensible Öffnungsprozesse. Sie forscht zur Transnationalisierung der feministischen Bewegung in Lateinamerika, den in Argentinien entstandenen Massendemonstrationen gegen Femizide *Ni Una Menos* und der Kampagne für das Recht auf legale, sichere und kostenlose Abtreibung, die durch das grüne Tuch repräsentiert wird.

Milena Pustet (B.Sc Informatik) ist wissenschaftliche Mitarbeiterin an der Hochschule für Technik und Wirtschaft (HTW) Berlin. Mithilfe von Techniken aus dem Bereich Data Science, Natural Language Processing und Machine Learning versucht sie, soziale Phänomene besser zu verstehen. In ihren aktuellen Studien konzentriert sie sich auf antisemitische und verschwörungstheoretische Diskurse in den sozialen Medien.

Elisabeth Steffen ist Informatikerin und Kulturanthropologin und forscht an den Schnittstellen von Machine Learning, Künstlicher Intelligenz und Political Data Science zu Verschwörungsideologien und Antisemitismus in sozialen Medien.

Illustrationen

Hamed Eshrat wurde 1979 in Teheran geboren, studierte visuelle Kommunikation an der Weißensee Kunsthochschule Berlin und an der Massey Universität in Wellington, Neuseeland. In seiner ersten Graphic Novel Tipping Point - Téhéran 1979 (2009 in Frankreich bei Edition Sarbacane erschienen) erzählt er die Geschichte seiner Familie während des politischen Umbruchs im Iran der 1970er-Jahre. Eshrats Deutschland-Debüt Venustransit spielt in Berlin, ist 2015 im avant-verlag erschienen und unter die zehn Finalisten des Deutschen Comicbuchpreisesder Berthold Leibinger Stiftung gewählt worden. 2018 folgt nach einem Szenario des Historikers Jochen Voit der Geschichtscomic Nieder mit Hitler!. Vier Jahre später ist mit Coming of H Eshrats dritter Comic im avant-verlag erschienen, eine autobiographische Geschichte, die vom Aufwachsen in der deutschen Provinz erzählt und ebenfalls unter den Finalisten des Comicbuchpreises gewählt wurde. Neben dem Comiczeichnen arbeitet Eshrat als Designer, freischaffender Künstler und Autor in Berlin. eshrat.de

Kooperations-partner

Haus der Kulturen der Welt

Inmitten tiefgreifender globaler und planetarer Transtormationsprozesse erkundet das HKW Künstlerische Positionen, wissenschaftliche Konzepte und politische Handlungsfelder neu. Es entwickelt und inszeniert ein in Europa einzigartiges Programm in einer Verbindung aus Diskurs, Ausstellungen, Konzerten und Performance, aus Forschung, Vermittlungsangeboten und Publikationen. Gemeinsam mit Künstler*innen, Wissenschaftler*innen, Expert*innen des Alltags und Partner*innen weltweit erkundet es Ideen, die im Entstehen begriffen sind, und teilt diese mit dem internationalen Publikum Berlins und der digitalen Offentlichkeit. Von 2019-2022 steht die Arbeit des HKW unter dem Titel „Das Neue Alphabet, worin vor allen Dingen Algorithmen als einflussreicen Gestalter der Gegenwart untersucht“ werden.

AMADEU ANTONIO STIFTUNG

Seit ihrer Gründung 1998 ist es das Ziel der Amadeu Antonio Stiftung, eine demokratische Zivilgesellschaft zu stärken, die sich konsequent gegen Rechtsextremismus, Rassismus und Antisemitismus wendet. Dafür unterstütz sie Initiativen und Projekte, die sich kontinuierlich für eine demokratische Kultur engagieren und Schutz von Minderheiten eintreten.

korientation

korientation ist eine (post)migrantische Selbstorganisation und ein kultur- und bildungspolitisches Netzwerk von Asiatischen Deutschen und Asiat*innen mit dem Lebensschwerpunkt Deutschland. korientation verfolgt das Ziel, den vielfältigen Lebenswirklichkeiten von Asiatischen Deutschen Präsenz und Ausdruck zu verleihen und sie damit bewusst und sichtbarer zu machen. Die Stärkung der kulturellen und politischen Selbstpräsentation von Asiatischen Deutschen ist maßgeblich, um einen Einfluss auf die Entwicklung einer vielfältigen deutschen Gesellschaft nehmen zu können. Dabei werden durch unterschiedliche Aktivitäten und Projekte rassistische Klischees und ausgrenzende Praktiken in Frage gestellt und der Blick für die unterschiedlichen Lebenswirklichkeiten von BPoC ausgeweitet. korientation arbeitet an der Schnittstelle von Kultur, Medien und Wissenschaft. Die Intersektion von politischem Aktivismus und wissenschaftlicher Forschung weiter auszuloten ist eines der Anliegen des Vereins.

ReachOut

ReachOut ist eine Beratungsstelle für Opfer rechter, rassistischer und antisemitischer Gewalt in Berlin. Die Situation und die Perspektive der Opfer rassistischer, rechter und antisemitischer Gewalt stehen im Zentrum der Arbeit. ReachOut bietet daneben antirassistische, interkulturelle Bildungsprogramme an und recherchiert rechtsextreme, rassistische und antisemitische Angriffe in Berlin und veröffentlicht dazu eine Chronik.

Die Vielfalt der Kulturen und Lebensweisen machen die besondere Attraktivität der Metropole Berlin aus. Demokratische Stadtkulturen und ein wertschätzendes Miteinander tragen diese Vielfalt und es gilt, sich immer wieder für sie einzusetzen. Gleichermaßen gilt es, das Recht aller Menschen auf Gleichbehandlung und Nichtdiskriminierung durchzusetzen. Eben dafür hat der Senat die Landesstelle für Gleichbehandlung – gegen Diskriminierung eingerichtet.

Die Kreuzberger Initiative gegen Antisemitismus - KIgA e.V. ist ein Träger der politischen Bildung, der innovative Konzepte für die pädagogische Auseinandersetzung mit Antisemitismus in der Migrationsgesellschaft als Ganzes entwickelt. Seit fast zwei Jahrzehnten erarbeiten sie modellhafte und lebensweltlich orientierte pädagogische Ansätze und Materialien für die politische Bildung und setzen diese in die Praxis um. Sie thematisieren komplexe, sensible und politisch brisante Themeninhalte und unterstützen mit unseren Kompetenzen Interessierte aus Bildung, Politik und Zivilgesellschaft. Sie qualifizieren bundesweit Multiplikator*innen, gestalten wissenschaftliche Diskurse aktiv mit und bieten Expertisen und Beratung für den Bildungsbereich, Politik und Gesellschaft.

Die Birds on Mars GmbH wurde 2018 in Berlin gegründet und hat in den letzten Jahren branchenübergreifend eine Vielzahl von Digitalisierungsprojekten mit Schwerpunkt auf Daten und Künstliche Intelligenz (KI) durchgeführt. Als cross-funktionales Team bestehend aus Daten- und KI-Expert*innen, Strateg*innen, Intelligence-Architects, Software-Entwickler*innen und Kreativen* unterstützt Birds on Mars Organisationen bei der Entwicklung von Strategien, Strukturen, Teams und Applikationen an den Verbindungslinien aus menschlicher und künstlicher Intelligenz. Zu den Kunden von Birds on Mars gehören u.a. die Deutsche Bahn, Teufel, Lufthansa, Diakonie und Stabilo. Mit ihrem firmeninternen Projekt sol haben sie einen offenen Raum geschaffen für Innovation, Denken und Erleben in Verbindung mit KI. In der Vergangenheit haben sie beispielsweise gemeinsam mit Künstler*innen die Artificial Muse entwickelt, mit welcher sie den Diskurs um KI und Kunst aktiv mitgestalten. In einem Projekt mit der Diakonie Rosenheim entwickeln sie KI für Kinder, und mit den Musiker*innen von Mouse on Mars erzeugen sie neue Töne und Klänge mit Hilfe von KI.

Das Bildungsteam Berlin-Brandenburg e.V. ist in der politischen Bildungsarbeit tätig und organisiert seit 1997 erfolgreich Seminare, Projekttage und Fortbildungen in Berlin, Brandenburg und darüber hinaus. Zum Team gehören sowohl ausgebildete Pädagog*innen als auch Politik- und Sozialwissenschaftler*innen. Sie verfügen über langjährige Erfahrungen in der außerschulischen Jugend- und Erwachsenenbildung. 2002 gründete das Bildungsteam Berlin-Brandenburg den Arbeitskreis „BildungsBausteine“ gegen Antisemitismus, der sich speziell mit dem Thema Antisemitismus befasst.

Mad About Pandas ist ein preisgekröntes Studio für Spiele- und kreative Medienproduktion, das von Patrick Rau, einem Spin-off der kunst-stoff GmbH mit Sitz in Berlin, gegründet wurde. Sie haben Expertise in Bereichen Interaktive Spiele, Medien- und Anwendungsproduktion für alle Arten von Publikum und Märkten, mit einem einzigartigen Gameplay und einem hohen konzeptionellen und künstlerischen Wert. Ihre Schwerpunkte liegen an Gameplay, Storytelling und Kunstdesign. Sie kombinieren ihre Kernkompetenzen aus allen Bereichen der interaktiven Medien und Plattformen im eigenen Haus. Sie gestalten emotionale Marken und entwerfen storyorientierte Inhalte mit ausgezeichnetem Design und masgeschneiderten Losungen. Durch Cross- und Transmedia-Konzepte stimmen sie alle Platformen und Zielgruppen aufeinander ab. Darüber hinaus nutzen sie Designworkshops zur Erstellung digitaler Kommunikationskonzepte und Workshops zur Wissenserweiterung.

DIGITALER HASS

Digitale Hassreden und Verschwörungsideologien in Zeiten der COVID-19 Pandemie

ASH Berlin Alice Salomon Hochschule Berlin University of Applied Sciences

htw Hochschule für Technik und Wirtschaft Berlin University of Applied Sciences

HKW Haus der Kulturen der Welt