



Wenn ethische Werte nur ein Feigenblatt sind

«Rassismus und Sexismus sind bei Google allgegenwärtig», sagt AI-Forscherin Timnit Gebru. Sie wollte dies ändern – und wurde entlassen, weil sie vor Diskriminierung durch Algorithmen gewarnt hat. «Do not feed the Google», Folge 4.

Von [Daniel Ryser](#), Ramona Sprenger (Text) und Adrià Fruitós (Illustration), 20.01.2023

Wir dürfen nicht predigen, ihr müsst in Mathematik, Informatik, Naturwissenschaften und Technik (Mint) gehen, wir müssen mit guten Beispielen kommen. Wir brauchen junge Frauen, die sagen: Ich mache eine Mint-Ausbildung, weil ich möchte beim Klima, bei der Nachhaltigkeit was tun (...) Solchen jungen Frauen müssen wir Bühnen geben, damit sie andere junge Frauen und Männer inspirieren.

Patrick Warnking, Country Director Google Switzerland, «Silicon Limmattal?», April 2022.

Die Worte von Google-Schweiz-Chef Patrick Warnking, einem weissen Mann in seinen Fünzigern, klangen hoffnungsvoll: junge Frauen, die bei Google arbeiten, um sich für das Klima einzusetzen. Es ist eine schöne Geschichte, und sie blieb an jenem Abend im Frühling auch unwidersprochen. Wir hatten aber die Worte von Bianca Wylie im Ohr, der Technologie-Aktivistin aus Kanada. Es sei beeindruckend, hatte Wylie zu uns gesagt, «wie Google jedes öffentliche Gut, jedes positive Narrativ als Waffe für seine Zwecke» einsetze.

Wir erreichen Timnit Gebru an ihrem Wohnort im Silicon Valley via Zoom vor ihrem Morgenkaffee. «Ich hoffe, es ist okay, wenn ich während des Gesprächs meine Hunde füttere», sagt die Computerwissenschaftlerin, eine der führenden Forscherinnen im Bereich der künstlichen Intelligenz. Sie wurde kürzlich vom Magazin «Time» in die Liste der 100 wichtigsten Persönlichkeiten 2022 aufgenommen. In der Laudatio hiess es: «Es braucht Mut, den mächtigsten Techkonzernen der Welt die Wahrheit ins Gesicht zu sagen. Und das ist es, was Timnit Gebru getan hat.»

Und weiter: «Gebru war die höchstrangige schwarze Frau und führte ein Team von Ethikerinnen zum Thema künstliche Intelligenz (KI) bei Google, angeworben, um Probleme zu eruieren und die Technologie zu verbessern. Sie wurde entlassen, nachdem sie ein Paper mitverfasst hatte, das genau das tat. Es legte rassistische Diskriminierung und Umweltbeeinträchtigungen innerhalb von gross angelegten Künstliche-Intelligenz-Systemen innerhalb des Unternehmens offen.»

Wir fragen Timnit Gebru, was sie davon halte, wenn der Chef von Google Schweiz bei einem Panel davon spreche, dass für Google Diversität, Frauenförderung und Nachhaltigkeit essenziell seien.

«*Oh, Bullshit!*», bricht es aus der Forscherin heraus, noch vor dem ersten Kaffee, und spätestens jetzt sind alle im Zoom-Call hellwach.

«Wisst ihr, es gibt diesen Begriff *diversity branding*. Und das ist genau das, was hier passiert», sagt Gebru. «Männern wie ihm bringt man bei: Sprecht über Diversität. Das ist das, was die Leute hören wollen.»

Technologie ist nicht neutral

Denn wenn man über Diversität spreche, gebe man den Leuten das Gefühl, dass Diskriminierung nicht existiere, auch wenn die Daten das Gegenteil zeigten.

«Bei Google ist Rassismus und Sexismus allgegenwärtig. Uns erreichten über unsere Google-Brain-Frauen-Mailingliste Nachrichten von überall auf der Welt. Aus Zürich hiess es: Man sehe ab einer gewissen Hierarchiestufe keine einzige Frau. Nur Männer. Meilenweit nur Männer und Männer.»

Nicht nur, dass Google die Forscherin Ende 2020 fristlos feuerte, nachdem sie sich geweigert hatte, ihr Paper zurückzuziehen – das Unternehmen be-

hauptete auch noch, dass Gebru selbst gekündigt habe. (Innert Stunden war eine Protestpetition lanciert, die tausendfach unterschrieben wurde, und sogar Mitglieder des US-Kongresses verlangten eine Erklärung von Google bezüglich Gebrus Entlassung.)

2019, während sie bei Google gemeinsam mit Margaret Mitchell das AI-Ethics-Laboratorium führte, hatte Gebru sich öffentlich dafür ausgesprochen, dass Amazon den Verkauf der Software zur Gesichtserkennung an Polizeibehörden stoppen müsse, weil die Technologie eben gar nicht neutral, sondern aufgrund der Daten voreingenommen sei gegenüber Frauen und People of Color (PoC); dass die Verwendung solcher Software letztlich zu gefährlich sei, um von Polizeibehörden verwendet zu werden.

Gebru war nach ihrem Studium in Stanford beseelt davon, Ethik in das Feld der künstlichen Intelligenz zu bringen, eines der lukrativsten Geschäftsfelder des 21. Jahrhunderts; ein Feld, das dominiert wird von weissen Männern. Dem «Time»-Magazin sagte Gebru, sie sei während ihrer Karriere entscheidend beeinflusst worden von einer Pro-Publica-Recherche zum Thema *predictive policing* aus dem Jahre 2016, die dargelegt habe, wie Gerichtssäle überall in den USA Software übernommen hatten, die vorgab, die Rückfallgefahr von Angeklagten voraussagen zu können – und somit die Richterinnen beim Strafmass entscheidend beeinflusste.



«Ich habe bei Google vor allem weisse Männer verärgert, die dachten, sie könnten mir sagen, was ich als Forscherin zu tun habe»: Timnit Gebru, AI-Forscherin. Cody O'Loughlin/The New York Times/laif

Pro Publica hatte die tatsächlichen Rückfallraten mit den Vorhersagen der Software verglichen und dabei herausgefunden, dass die Software nicht nur häufig falschlag, sondern auch gefährlich und rassistisch voreingenommen war: «Die Software stufte viel eher schwarze Angeklagte, die dann gar nicht rückfällig wurden, als Personen mit hohem Rückfallrisiko ein. Und sie stufte wiederum weisse Angeklagte, die dann tatsächlich rückfällig wurden,

als Personen mit tiefem Rückfallrisiko ein», heisst es im «Time»-Porträt über Gebru. Die Pro-Publica-Resultate hätten gezeigt, wie ein KI-System aufgrund historischer Daten trainiert wird, welche die Ungleichheiten widerspiegeln; dieses System projiziere solche Ungleichheiten dann automatisiert und skaliert in die Zukunft.

Als Person of Color, die die eigenen Erfahrungen mit der Polizei reflektiere, hätten sie diese Ergebnisse in Angst und Schrecken versetzt. Denn niemand, der ihre Erfahrungen teilte, sei in die Entwicklung dieser Programme involviert. Das Feld der künstlichen Intelligenz, bei Google und anderswo, sei nicht divers besetzt, sondern von weissen Männern, und reproduziere schliesslich bestehende Machtverhältnisse und Rassismus. «Ich bin sehr besorgt über die Zukunft der künstlichen Intelligenz», zitierte das «Time»-Magazin Timnit Gebru. «Nicht wegen der Gefahr, dass bösartige Maschinen die Macht übernehmen, sondern wegen der homogenen eindimensionalen Gruppe von Männern, die derzeit an der Weiterentwicklung der Technologie beteiligt sind.»

Ständig sei man dabei, mit Vorschriften, Verboten, Regulierungen den Schlamassel aufzuräumen, den die CEOs von Big Tech anrichteten, «während die schon dabei sind, die Zukunft zu gestalten, über das Metaverse reden oder was auch immer». Das müsse sich ändern, sagt Gebru: «Wir anderen müssen darüber nachdenken, welche Zukunft wir wollen. Für die Entwicklung von künstlicher Intelligenz würde eine hoffnungsvolle Zukunft bedeuten, dass man bei der Entwicklung die Bedürfnisse jener Menschen ins Auge fasst, die in der Gesellschaft am meisten benachteiligt sind.»

«Giftcocktail für die Demokratie»

Ruha Benjamin, eine Soziologin und Professorin am Department of African American Studies der Princeton-Universität, beschreibt diese Vorgänge in ihrem Buch «Race After Technology». Mit Blick auf Algorithmen und Vorhersagemodelle schreibt Benjamin von einem «Einsatz neuer Technologien, welche bestehende Ungleichheiten widerspiegeln und reproduzieren, aber als objektiver oder fortschrittlicher als die diskriminierenden Systeme einer früheren Ära beworben und wahrgenommen werden».

Im Buch «Algorithms of Oppression» von Safiya Noble, Professorin an der University of California und Co-Direktorin des Center for Critical Internet Inquiry, stecken mehr als sechs Jahre akademische Forschung zu Google-Suchalgorithmen. Noble argumentiert, die Google-Suchalgorithmen seien nicht neutral, sondern würden strukturelle Diskriminierung aufrechterhalten.

Sie wurde auf das Phänomen aufmerksam, als sie 2011 «schwarze Mädchen» googelte, um Freizeitangebote für ihre Tochter und ihre kleinen Nichten zu finden. Die Suchmaschine spuckte rassistische Pornos aus. Das Ergebnis ihrer Forschung: Die Google-Suchalgorithmen werden mit sexistischem und rassistischem Angebot im Netz gefüttert und lernen von den Nutzern, dass dieses Angebot eine grosse Nachfrage hat. Und reproduzieren dann die vorherrschenden Stereotype und Vorurteile.

Google man Namen, würden in der Suchmaschine bei «schwarzen Namen», wie die afroamerikanische Informatikerin Latanya Sweeney schreibt, vornehmlich Verhaftungsprotokolle auftauchen, bei klassisch «weissen Namen» hingegen Anwaltskanzleien und Arztbüros. «Googles Algorithmen optimierten sich für die rassistischen Muster früherer Nutzer, die auf diese

Anzeigen geklickt hatten, lernten dann die rassistischen Vorlieben einiger Nutzer und gaben sie an alle anderen weiter», schreibt Sweeney.

Die Verstärkung von Ungleichheit treffe nicht nur in Bezug auf *race* oder Geschlecht zu, sondern auch in Bezug auf Klasse, wie verschiedene wissenschaftliche Bücher darlegen, so zum Beispiel der «New York Times»-Bestseller «Weapons of Math Destruction» der Mathematikerin Cathy O’Neil.

Es sei «besonders beunruhigend», schreibt O’Neil, wie die Algorithmen auch die Diskriminierung von Armutsbetroffenen verstärkten. Sie schreibt von einem «Giftcocktail für die Demokratie»: «Wenn eine arme Studentin keinen Kredit bekommt, weil ein Kreditvergabemodell sie wegen ihrer Postleitzahl als zu riskant einstuft, wird sie von der Art von Bildung abgeschnitten, die sie aus der Armut herausführen könnte, und es entsteht ein Teufelskreis. Die Modelle stützen die Glücklichen und bestrafen die Unterdrückten.»

Die Wut der weissen Männer

Zahlreiche Kolleginnen warnten Gebru vor der Arbeit in einer Firma, bei der die Zahl von PoC bei den Vollzeitangestellten im einstelligen Prozentbereich liegt. Gebru sollte die erste schwarze Frau in der gesamten Forschungsabteilung werden. Und dass jemand wie sie – renommiert, starrsinnig, kritisch – für das Unternehmen arbeitete, wurde gerne als Beweis dafür gewertet, dass Techkonzerne durchaus an den Risiken ihrer Arbeit interessiert seien, ja sich womöglich sogar selber regulieren könnten.

Im Paper, das schliesslich Ende 2020 zu ihrer Entlassung bei Google führte, warnten Gebru und ihr AI-Ethik-Team vor übermässigem Enthusiasmus bezüglich *large language models* und nannten diverse Risiken, die man in Bezug auf diese Technologie unbedingt in den Griff bekommen müsse. Dabei handelt es sich um KI-Modelle, die mit unzähligen Textdaten gefüttert und trainiert werden, mit dem Ziel, selbst Texte generieren oder gar interpretieren zu können.

Zu den Problemen, die Gebru im Paper identifizierte, gehörten unter anderem ein hoher Energieverbrauch und dessen Folgen für die Umwelt, sexistische oder rassistische Textmuster, die sich durch die Trainingsdaten in die Technologie schleichen können und dann reproduziert werden; aber auch die Gefahr von missbräuchlicher Anwendung; beispielsweise im Herstellen von Falschinformationen.

«Das Paper, wofür ich schliesslich entlassen wurde, war nur die Spitze des Eisbergs», sagt Timnit Gebru beim Morgenkaffee. «Ich war beim Management und bei der Personalabteilung von Anfang an nicht beliebt.» Man habe sie zu Google geholt, damit sie Probleme adressiere und Verbesserungsvorschläge mache. Also habe sie dem Management immer wieder und wieder geschrieben, was man ändern müsste. Das Management habe sie zu Meetings getroffen, ein ums andere Mal, um ihr das Gefühl zu geben, gehört zu werden, ihr das Gefühl zu geben, man tue ihr damit einen Gefallen.

«Aber nichts ist passiert. Überhaupt nichts. Der Fakt, dass ich die Dreistigkeit hatte, darauf zu bestehen, dass man mich nicht nur anhörte, sondern dass meine Erkenntnisse auch Folgen hätten – das war der Grund für meine Absetzung», sagt Gebru. «Dass ich die Frechheit hatte, auf meiner wissenschaftlichen Integrität zu bestehen und darauf, meine Erkenntnisse auch zu publizieren und meine Forschung nicht an die Vorgaben anzupassen. Bei Google glaubte man, man könne meine Integrität untergraben und mir sa-

gen, was ich zu publizieren habe», sagt die renommierte Forscherin. «Aber ich wollte meine wissenschaftliche Integrität schützen. Ich habe vor allem weisse Männer verärgert, die dachten, sie könnten mir sagen, was ich als Forscherin zu tun habe», sagt die Afroamerikanerin.

«Google betreibt *ethics washing*»

Für ihr Ethical-AI-Team bei Google hatte Gebru ein diverses und interdisziplinäres Team rekrutiert aus Soziologinnen und Anthropologen. Alex Hanna, die als eine der ersten Sozialwissenschaftlerinnen ins Team rekrutiert worden war, sagt gegenüber der Republik: «Was Gebrus Arbeit von jener vieler anderer unterscheidet, ist ihre Fähigkeit, über technische Kritik hinauszugehen.» Während die meisten Ingenieure sich auf technische Probleme konzentrierten, habe sie stets die grösseren Themen im Blick. «Statt zu fragen, wie man Datensätze verbessert, fragt sie: Für welche Zwecke werden diese gebraucht? Wer hat sie entwickelt und mit welchem Ziel? Und wollen wir diese überhaupt?»

Im Unterschied zu vielen anderen habe Gebru von Anfang an grossen Wert auf Interdisziplinarität gelegt und stets die Menschen ins Zentrum gestellt, «besonders diejenigen, die keine eigene Lobby haben». Hanna weiss, wovon sie spricht. Die promovierte Soziologin ist auch eine LGBTQIA+-Aktivistin, die sich für die Rechte von trans Personen einsetzt und Datensätze auf Geschlechtsstereotype untersucht. Ohne Gebru, sagt sie, wäre sie kaum bei einem Technologieunternehmen wie Google gelandet.

Gebru spricht heute von «Gaslighting»: «Gegen aussen verkündete Google, man sponsere Grace Hopper Celebration, die grösste Konferenz für Frauen in Computing. Am selben Tag schickten mir dieselben Leute E-Mails, in denen sie Vergeltung an mir übten für meine Arbeit.» Google habe Grace Hopper bewusst benutzt als Gaslighting-Mechanismus. Und ebendieses Gaslighting – wie bei Google einerseits öffentlich Grace Hopper gesponsert und gleichzeitig intern die afroamerikanische Co-Leiterin des Ethik-AI-Labors fertiggemacht und rassistisch attackiert wurde – legte Gebru in einer Rede ausführlich dar (unter anderem, indem sie E-Mails mit Anfeindungen vorlas, die sie von weissen Google-Kollegen erhalten hatte, als sie 2021 Keynote-Speakerin ebenjener Grace-Hopper-Konferenz war).

«Hier ist der Punkt: Kennt ihr den Begriff *ethics washing*? Das ist es, was Google macht. Es ist ähnlich wie Greenwashing», sagt Gebru. Unternehmen würden die Arbeit von Menschen wie ihr benutzen, um der Welt zu zeigen, dass man viel mache im Bereich Ethik und künstliche Intelligenz. Dass man Fragen und Bedenken ernst nehme.

«*Ethics washing* ist der Grund, warum Google mich angestellt hat. Es ist ein Verteidigungsmechanismus», sagt Gebru. «Damit man sagen kann: «Wir haben all diese Papers, wir zensieren nicht. Der Staat muss uns nicht regulieren. Wir regulieren uns selbst.» So nutzen solche Unternehmen die Arbeit, die ich mache. Denn es ist ja nicht nur so, dass ich kein Geld einbringe, sondern dass ich auch noch sage: «Anstatt dass wir jetzt 100 Dollar verdienen, verdienen wir lieber nur 10 Dollar. Wir sollten verlangsamen. Wir sollten mehr Ressourcen in Sicherheit stecken.»»

Big Tech habe viel unternommen, die Geschichte ins öffentliche Bewusstsein zu pflanzen, dass man sich selber reguliere, dass es keine staatlichen Regulationsmechanismen brauche. Deswegen sei das *ethics washing* so wichtig. «Dass sich ein Unternehmen, das derart nach Profit strebt, selber reguliert – diese Vorstellung ist geradezu absurd», sagt Gebru. «Bei wel-

cher anderen Industrie würde man davon ausgehen, dass sie sich selbst reguliert? Dass es ohne Vorschriften geht, dass ein Unternehmen schon selber schaut, dass es etwa Vorschriften im Bereich der Umweltverschmutzung einhält? Oder bei der Computer-Hardware: wie viel Strahlung ein Gerät verursachen kann? Dass Big Tech so lange nicht reguliert war und es grösstenteils noch immer nicht ist, ergibt keinen Sinn.»

«Nie da gewesene Konsolidierung von Macht»

Wir fragen die ehemalige Co-Leiterin der Abteilung für Ethik in der künstlichen Intelligenz, warum sich die Menschen in der Schweiz, in Zürich, die lokale Politik, die öffentlichen Einrichtungen um Google Gedanken machen sollten. Warum wir uns kümmern sollten. Warum wir uns nicht von schönen Worten über Nachhaltigkeit und Diversität abspesen lassen sollten.

«Das Bedeutendste, was im Silicon Valley momentan passiert, das ist nicht Disruption, also Erschütterung, Umwälzung», sagt Timnit Gebru. «Das Bedeutendste, was derzeit im Silicon Valley passiert, ist die Konsolidierung von Macht. Und zwar in einer Geschwindigkeit und in einer Grössenordnung, wie es sie in der Geschichte wahrscheinlich noch nie gegeben hat.»

Wenn ein multinationaler Konzern in der Nachbarschaft einen riesigen Digital Hub baue, dann müsse die Öffentlichkeit wissen, was dieser Konzern tue. «Die Öffentlichkeit soll in der Lage sein, diesen Konzern zu überprüfen. Sie soll wissen, wie er Daten sammelt und welche Daten er anwendet. Deswegen sollen die Leute in Zürich sich mit Google beschäftigen. Mit dieser ungesesehenen Zentralisierung von Macht und der damit einhergehenden umfassenden Kontrolle unserer Leben», sagt Gebru.

Wenn die ganze Welt dieselbe App benutze, beispielsweise Youtube: dann zeige sich, dass das nicht einfach ein grosses Unternehmen sei, sagt Gebru. «Google ist ein weltpolitischer Akteur.» Es gebe viel Forschung dazu, wie Big Tech und Google die Weltpolitik beeinflussten.

«Wenn öffentliche Institutionen und Unternehmen sich auf Big Tech einlassen, geht es schnell nicht mehr um das öffentliche Interesse, sondern nur noch um den Profit des Konzerns», sagt Gebru. Alles werde für den Profit einiger weniger Unternehmen getan.

Öffentliche akademische Einrichtungen würden von Google für ihre Forschungsarbeiten bezahlt. Google-Lobbyisten beeinflussten die Regierungen und die Gesetzgebung. Ob es um Gesetze gehe bezüglich Klimawandel oder künstlicher Intelligenz. «Die Leute müssen ihre öffentlichen Institutionen schützen, die sie mit ihren Steuergeldern finanzieren», sagt Gebru. «Ansonsten geraten wir in eine Situation, in der die *checks and balances* verloren gehen und alles für den Profit einiger weniger Unternehmen getan wird.»

«Die Situation ist dramatisch»

Dublin, sagt Timnit Gebru am Ende unseres Gesprächs, sei ein Dreh- und Angelpunkt für Techkonzerne, weil dort diesen Unternehmen massive Steuererleichterungen gewährt würden.

Die Obdachlosigkeit sei daraufhin in der Stadt explodiert. Die Leute seien verdrängt worden, weil sie es sich nicht mehr leisten könnten, in Dublin zu leben.

«Nehmen wir als anderes Beispiel die Bay Area in Kalifornien, wo diese Konzerne daheim sind und wo ich lebe», sagt Gebru. «Die Situation ist dramatisch. Und zwar wegen Firmen wie Google, die gleichzeitig davon reden, die Welt zu retten.»

Wenn die Leute irgendwo auf der Welt und vermutlich auch in Zürich an das Silicon Valley denken würden, dann denke niemand daran, dass die Obdachlosigkeit in der Bay Area wegen Big Tech explodiert sei, sagt Gebru.

«Big Tech zahlt so hohe Löhne, dass die Mieten in San Francisco für normale Leute, für Lehrerinnen beispielsweise, unerschwinglich geworden sind. Wenn ich schaue, was ich und meine Kolleginnen bei Google verdienen haben: Niemand sollte so viel Geld verdienen», sagt Gebru. «Und dann hast du Menschen, die arbeiten zu Tiefstlöhnen in der Google-Cafeteria. Mit diesem Lohn können sich diese Menschen die Mieten, die Google verursacht, nicht mehr leisten. Sie arbeiten in der Google-Cafeteria und sind gleichzeitig obdachlos oder leben in ihrem Auto.»

Zur Serie «Do not feed the Google» und zur Co-Autorin

Diese Serie ist eine Zusammenarbeit zwischen der Republik, dem Dezentrum und dem WAV. Das Dezentrum ist ein Think & Do Tank für Digitalisierung und Gesellschaft. Hinter dem Dezentrum steht ein öffentlicher, gemeinnütziger Verein. Das WAV ist ein unabhängiges Recherchekollektiv aus Zürich.

Ramona Sprenger ist Interaction Designerin aus Zürich. Als Partnerin beim Think & Do Tank Dezentrum engagiert sie sich für eine nachhaltige Digitalisierung, bei der die Gesellschaft im Mittelpunkt steht. Aktuell arbeitet sie für TA-Swiss an einer Publikation zu Blockchain und Kultur und baut mit Climate Ticker eine Plattform für offene, lokale Klimadaten und lokalpolitische Massnahmen auf.