

MEINUNGSBAROMETER.INFO

DAS FACHDEBATTENPORTAL

DOKUMENTATION

FACHDEBATTE

KENNEN ALGORITHMEN GESCHLECHTER?

Wie KI von Geschlechterrollen-Stereotypen in die Irre geführt wird

Die Dokumentation beinhaltet alle Positionen, ausführliche Analysen und Prognosen zu dieser Fachdebatte sowie eine übersichtliche Management Summary.

Debattenlaufzeit: 23.03.2023 - 15.09.2023

INHALTSVERZEICHNIS

DEBATTENBESCHREIBUNG	4
AKTIVE DEBATTENTEILNEHMER	5
MANAGEMENT SUMMARY	7
DEBATTENBEITRÄGE	10
WARUM GESCHLECHTERROLLENSTEREOTYPE OFFENGELEGT WERDEN SOLLTEN Und warum Datensätze geprüft und an die Realität angepasst werden müssen	10
PROF. DR. SWETLANA FRANKEN Leiterin der Denkfabrik Digitalisierte Arbeitswelt FH Bielefeld, Fachbereich Wirtschaft	
DEN DIGITALEN TRANSFORMATIONSPROZESS MIT DEM VERFASSUNGSaufTRAG IN EINKLANG BRINGEN Wie die Digitalisierung gendergerecht umgesetzt werden kann	13
DR. STEFAN ULLRICH Stellvertretender Sprecher des Fachbereichs Informatik und Gesellschaft Gesellschaft für Informatik (GI)	
DISKRIMINIERUNGEN WIRKSAM BEGEGNEN Geschlechter müssen für Datenerfassung und -verarbeitung sichtbar sein	16
PROF. DR. NICOLA MARSDEN Forschungsprofessur für Sozioinformatik, stellvertretende Vorsitzende des Kompetenzzentrums Technik-Diversity-Chancengleichheit Hochschule Heilbronn	
WO TECHNOLOGINNEN UND TECHNOLOGEN IHRE GRENZEN KENNEN Wie man Künstliche Intelligenz geschlechtergerecht machen kann	20
DR. ALJOSCHA BURCHARDT Principal Researcher Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) Berlin	
WIE KI-ANWENDUNGEN STEREOTYPEN REPRODUZIEREN Und warum auch Algorithmen gendersensitiv sein müssen	23
NADIA FISCHER CEO & Co-Founder Witty Works	

ETHIK IST NICHT DIE FAHRRADBREMSE AM INTERKONTINENTALFLUGZEUG Warum eine ethische Perspektive notwendig ist, um Diskriminierungsmechanismen offenzulegen	30
PROF. DR. REGINA AMMICHT-QUINN Sprecherin IZEW Universität Tübingen	
ÜBER DIE ILLUSION VON OBJEKTIVITÄT BEI DER NUTZUNG VON KI- SYSTEMEN Wie KI die Realität richtig erkennen kann	35
PROF. DR. SABINE T. KÖSZEGI Professorin TU Wien	
IMPRESSUM	38

DEBATTENBESCHREIBUNG



INITIATOR

UWE REMPE

Freier Journalist

Meinungsbarometer.info

KENNEN ALGORITHMEN GESCHLECHTER?

Wie KI von Geschlechterrollen-Stereotypen in die Irre geführt wird

Ist KI von Haus aus fair, unvoreingenommen und egalitär? Nein, sagen Forscher und Praktiker. So erkennt beispielsweise die Software in Zoom-Konferenzen und ähnlichen Diskussionsformaten laut einer Studie von Forschern der dänischen Universität Sonderborg weibliche Stimmen und Gesichter schlechter. Gängige Suchmaschinen, heißt es anderswo, würden männliche und weibliche Berufsbezeichnungen nicht als synonym erkennen und anzeigen, wobei oft die männliche Berufsbezeichnung in den Stellenanzeigen genutzt wird.

Etwas anders gelagert: Aus der Medizin ist das sogenannte Gender-Data-Gap bekannt, die Lücke also, die sich bei der Erhebung von Daten vor allem männlicher Studienteilnehmer (zum Beispiel Medikamententests, Diagnostik von typischen Krankheitssymptomen) ergibt, und deren Ergebnisse unterschiedslos für Frauen und Männern gültig sein sollen.

Das sind nur einige wenige Probleme, die sich in der Praxis in Sachen Gendergerechtigkeit auftun. Wie können in Zeiten der Digitalisierung jeder und jedem die Daten zugeordnet werden, die für sie/ihn relevant sind? Ist das eine Frage der Programmierung, von bewussten oder unbewussten Vorurteilen oder Prägungen der Ingenieure/Programmierer? Was ist zu tun, um KI bzw. Suchalgorithmen von vornherein geschlechtergerecht anzulegen?

AKTIVE DEBATTENTEILNEHMER



PROF. DR. REGINA AMMICHT-QUINN

Sprecherin IZEW
Universität Tübingen



DR. ALJOSCHA BURCHARDT

Principal Researcher
Deutsches Forschungszentrum für Künstliche
Intelligenz (DFKI) Berlin



NADIA FISCHER

CEO & Co-Founder
Witty Works



PROF. DR. SWETLANA FRANKEN

Leiterin der Denkfabrik Digitalisierte Arbeitswelt
FH Bielefeld, Fachbereich Wirtschaft



PROF. DR. SABINE T. KÖSZEGI

Professorin
TU Wien



PROF. DR. NICOLA MARSDEN

Forschungsprofessur für Sozioinformatik,
stellvertretende Vorsitzende des Kompetenzzentrums
Technik-Diversity-Chancengleichheit
Hochschule Heilbronn



DR. STEFAN ULLRICH

Stellvertretender Sprecher des Fachbereichs Informatik
und Gesellschaft
Gesellschaft für Informatik (GI)

MANAGEMENT SUMMARY

15.09.2023 | SUMMARY

WARUM KI GENDERSENSIBEL SEIN MUSS

Und welche Werte dabei eine große Rolle spielen



Uwe Rempke - Freier Redakteur [Quelle: Meinungsbarometer.info]

"Genderungerechte KI wird uns als Gesellschaft in die Vergangenheit katapultieren, wenn wir jetzt nicht handeln", sagt Swetlana Franken, Professorin für Betriebswirtschaftslehre an der Fachhochschule Bielefeld. Das Problem müsse durch vielfältige technische Maßnahmen und gesellschaftliches Bewusstsein angegangen werden. Das Problem: „Die durch KI und Algorithmen generierten Entscheidungen und Empfehlungen sind so gut, wie die zugrundeliegende Datenbasis, an der sie trainiert wurden. Solange in der Gesellschaft und Wirtschaft (und damit auch in den Daten) Geschlechterrollen-Stereotype verbreitet sind, wird auch die KI diese Stereotype widerspiegeln.“

Im Hinblick auf Gendergerechtigkeit im digitalen Raum braucht es für Dr. Stefan Ullrich von der Gesellschaft für Informatik (GI) "eine soziotechnische Perspektive auf die

Digitalisierung hinsichtlich Gestaltung, Nutzung und Zugang". Der Informatiker und Philosoph rät deshalb zu "kluger Gestaltung" der digitalen Transformation. Öffentliche Stellen hätten bereits Wege eröffnet, wie etwa die Umsetzungsstrategie „Digitalisierung gestalten“ und das Gutachten zum Dritten Gleichstellungsbericht der Bundesregierung zeigten. Als Richtschnur des Handelns diene dabei Artikel 3 Abs. 2 des Grundgesetzes. „Die Gesellschaft für Informatik gibt in ihrem Ethischen Kompass eine Orientierung.“

"Richtig gestaltet, erweitern KI-Systeme unsere Möglichkeiten, eine diskriminierungsfreie und geschlechtergerechtere Welt zu schaffen", sagt Prof. Dr. Nicola Marsden, Inhaberin der Forschungsprofessur für Sozioinformatik an der Hochschule Heilbronn und stellvertretende Vorsitzende des Kompetenzzentrums Technik-Diversity-Chancengleichheit. „Wir müssen sicherstellen, dass bei der Datenerfassung und -verarbeitung eine ausreichende Anzahl von Datenpunkten für alle Geschlechter vorhanden ist. Das heißt also paradoxerweise: Wenn wir wollen, dass Geschlecht keine Rolle mehr spielt, müssen wir ihm besondere Aufmerksamkeit widmen. Nur dann können wir existierende Ungerechtigkeiten korrigieren.“

Um Künstliche Intelligenz von vornherein gendergerecht aufzustellen, muss man zunächst einmal "ein Sensorium dafür haben, dass überhaupt eine Benachteiligung stattfinden kann", sagt Dr. Aljoscha Burchardt, Principal Researcher und stellvertretender Standortsprecher am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) in Berlin. Idealerweise sei man „Gender-neutral by Design“. Ein wichtiger Hebel seien die Trainingsdaten. „Die komplexen Antworten auf alle diese Fragen müssen wir in einem gesamtgesellschaftlichen Diskurs aushandeln.“

"Sprache ist nicht neutral. Wörter, die wir nutzen, um unsere Umwelt zu beschreiben, bestimmen, wie wir diese Welt sehen und empfinden", erklärt Nadia Fischer den Ausgangspunkt ihrer Unternehmung: Sie ist CEO & Co-founder des KI-basierten digitalen Schreibassistenten Witty, der in Wirtschaft und Gesellschaft für mehr Inklusivität sorgen soll. Sie bearbeitet das Problem vorrangig in Bezug auf "Large Language

Models" (LLM). Sie empfiehlt aktuell eine „Überprüfung der Resultate von Sprach-KI mit dafür gemachten Tools, die Voreingenommenheiten und Stereotypen in den LLM-Texten erkennen können. Diese Tools werden 'AI Whisperer' genannt. Sie flüstern den LLM-Texten gender-neutralere und unvoreingenommene Lösungen ein“.

"Wir brauchen mehr Frauen und Minderheiten im KI-Bereich, aber wir brauchen genauso ein größeres Bewusstsein für Diskriminierungsvorgänge bei allen, die in diesem Bereich arbeiten und bei allen, die die Wirtschaft dieser Bereiche beeinflussen", macht Prof. Dr. Regina Ammicht Quinn die Komplexität der Thematik deutlich. Sie ist unter anderem Sprecherin und Leiterin des Bereichs Gesellschaft, Kultur und technischer Wandel des Wissenschaftlichen Rats des Internationalen Zentrums für Ethik in den Wissenschaften (IZEW) der Universität Tübingen. Einfache Lösungen gibt es nicht, zeigt sie detailliert in ihrem Beitrag auf.

Geschlechterstereotypen schlagen sich auch in algorithmischen Systemen nieder - was zur Perpetuierung von Stereotypen und Vorurteilen führt, ungerechtfertigter Diskriminierung und Fehlentscheidungen, erklärt Prof. Dr. Sabine T. Köszegi vom Institute of Management Science der Technischen Universität Wien. „Mit der Bezeichnung ‚Algorithmus‘ wird verschleiert, dass kulturelle, gesellschaftliche und politische Werte lediglich in eine mathematische Sprache übersetzt sind und es entsteht eine Illusion von Objektivität bei der Nutzung von KI-Systemen, die nicht gerechtfertigt ist.“ Ihr Fazit: „Die Potenziale von KI-Technologie können nur dann gehoben werden, wenn sie gendersensibel und inklusiv, die Menschenrechte achtend gestaltet ist. Das ist ausschließlich eine Frage unserer Werte und keine Frage der technischen Machbarkeit.“

DEBATTENBEITRAG

25.04.2023 | INTERVIEW

WARUM GESCHLECHTERROLLENSTEREOTYPE OFFENGELEGT WERDEN SOLLTEN

Und warum Datensätze geprüft und an die Realität angepasst werden müssen



Prof. Dr. Swetlana Franken, lehrt an der FH Bielefeld Allgemeine Betriebswirtschaftslehre, Personalmanagement und Schlüsselqualifikationen des Managements [Quelle: FH Bielefeld]

"Genderungerechte KI wird uns als Gesellschaft in die Vergangenheit katapultieren, wenn wir jetzt nicht handeln", sagt Swetlana Franken, Professorin für Betriebswirtschaftslehre an der Fachhochschule Bielefeld. Das Problem müsse durch vielfältige technische Maßnahmen und gesellschaftliches Bewusstsein angegangen werden.

Inwieweit ist eine gendergerechte KI, sind intelligent programmierte und gendersensitive Algorithmen für Sie ein praktisch relevantes Problem?

Die durch KI und Algorithmen generierten Entscheidungen und Empfehlungen sind so

gut, wie die zugrundeliegende Datenbasis, an der sie trainiert wurden. Solange in der Gesellschaft und Wirtschaft (und damit auch in den Daten) Geschlechterrollen-Stereotype verbreitet sind, wird auch die KI diese Stereotype widerspiegeln. Auch wenn ChatGPT – nach seiner Geschlechtergerechtigkeit gefragt – behauptet, geschlechtergerecht zu sein, stützt es seine Aussagen auf die Internetdaten aus der Vergangenheit, die Stereotype und Vorurteile beinhalten. Diese Problematik wird dadurch verstärkt, dass KI meistens von Männern programmiert und genutzt wird, sodass die Sensibilität für die Verzerrungen kaum vorliegt.

In welchen Bereichen sehen Sie eine besondere Dringlichkeit für entsprechende Lösungen?

Beim Trainieren von Algorithmen sollte man die Datensätze auf geschlechterspezifische (und allgemeiner auf jegliche) Diskriminierungen überprüfen und gegebenenfalls korrigieren bzw. vervollständigen. Die Daten müssen eine ausgewogene und repräsentative Stichprobe aller Geschlechter enthalten. Dies ist besonders dringend im Recruiting, wo Algorithmen mittlerweile selbstständig nach geeigneten Kandidat(inn)en im Netz suchen. Da Männer als Führungskräfte eher präsent und erfolgreich sind, sucht KI in erster Linie nach Männern.

Was ist zu tun, damit Algorithmen die Geschlechter unterscheiden und adäquat zuordnen können? Welche Herausforderungen sind dabei zu lösen?

Mangelnde Daten (speziell zu Frauen) erschweren diese Prozesse. Was hilft? Daten vervollständigen bzw. bereinigen, geschlechtergemischte Teams für programmieren und trainieren einsetzen, die Ergebnisse durch Frauen testen lassen. Gefragt ist auch eine KI-Ethik für Informatiker(innen), die mithilfe technischer Lösungen wie DADM (Discrimination-Aware Data-Mining) oder FATML (Fairness, Accountability and Transparency in Machine Learning) zu einer diskriminierungsfreien KI beitragen soll.

Allerdings können nicht alle Probleme nur durch eine Erhöhung des Frauenanteils an Programmierenden, Entwickelnden, Testenden gelöst werden (auch wenn das an sich schon schwer genug ist!). Die Wurzeln der Geschlechterungerechtigkeit liegen viel tie-

fer. Die Vorurteile gegenüber Frauen sind in den Köpfen von Männern und Frauen (!) und bleiben unbewusst. Ein bedeutender Schritt besteht darin, die Geschlechterrollenstereotype offenzulegen, bei sich selbst und den Anderen zu erkennen, zum relevanten Gesprächsthema in der Gesellschaft, im Unternehmen, in der Familie zu machen. Ein Meeting um 17 Uhr? Ein Zukunftsworkshop am Wochenende? Nein, danke! Die Unternehmenskultur sollte frauen- und familienfreundlicher werden.

Ist Gendergerechtigkeit vorrangig eine öffentliche Aufgabe oder auch privatwirtschaftlich relevant? Welche rechtlichen Vorgaben sollten diesbezügliche Algorithmen bekommen?

Die KI-Gendervorurteile beeinflussen automatisierte Entscheidungen und Empfehlungen sowohl in der Privatwirtschaft als auch im öffentlichen Sektor, insofern ist das eine Aufgabe für alle. Auch die Hochschulen, mit einem sehr geringen Anteil von hochdotierten Professorinnen, sollten sich mit dem Problem intensiver befassen. Auch die UNESCO (2022) formuliert eine KI-Ethik-Empfehlung und schlägt Maßnahmen vor wie diversere Zusammensetzung von KI-Teams, AI Ethics Officer in Betrieben etc.

Schon heute ist KI eine bedeutende Querschnittstechnologie in vielen Lebensbereichen, Tendenz steigend. Genderungerechte KI wird uns als Gesellschaft in die Vergangenheit katapultieren, wenn wir jetzt nicht handeln.

DEBATTENBEITRAG

28.04.2023 | INTERVIEW

DEN DIGITALEN TRANSFORMATIONSPROZESS MIT DEM VERFASSUNGSaufTRAG IN EINKLANG BRINGEN

Wie die Digitalisierung gendergerecht umgesetzt werden kann



Dr. Stefan Ullrich - Stellvertretender Sprecher des Fachbereichs Informatik und Gesellschaft, Gesellschaft für Informatik (GI)
[Quelle: GI/Frl. von Phön]

Im Hinblick auf Gendergerechtigkeit im digitalen Raum braucht es für Dr. Stefan Ullrich von der Gesellschaft für Informatik (GI) "eine soziotechnische Perspektive auf die Digitalisierung hinsichtlich Gestaltung, Nutzung und Zugang". Der Informatiker und Philosoph rät deshalb zu "kluger Gestaltung" der digitalen Transformation.

Inwieweit ist eine gendergerechte KI, sind intelligent programmierte und gendersensitive Algorithmen für Sie ein praktisch relevantes Problem?

Bei der KI-gestützten automatischen Übersetzung von Texten in eine andere Sprache kommen regelmäßig schon für überholt geglaubte Gesellschaftsbilder zum Vorschein, aus Ärztinnen werden Ärzte, aus Erziehern werden Erzieherinnen. Wer immer

mit informationstechnischen Systemen herstellt oder auch nur mit ihnen umgeht, ist auf eine gendergerechte Technik insbesondere im Bereich der KI angewiesen, um keine unbeabsichtigte Diskriminierung zu erfahren oder zu verüben. Die Studie von Carsten Orwat zum Diskriminierungspotenzial von Algorithmen ist in diesem Zusammenhang sehr lesenswert.

In welchen Bereichen sehen Sie eine besondere Dringlichkeit für entsprechende Lösungen?

In Diskussionen um den Einsatz von Algorithmen für die Arbeitswelten der Zukunft wird deutlich, dass der voranschreitende Prozess der Digitalisierung kluger Gestaltung bedarf. Besonders in der Digitalbranche, aber auch der Wirtschaft und der Politik insgesamt ist die Dringlichkeit in besonderem Maße gegeben. Selbst wenn beispielsweise Geschäftsmodelle der datengetriebenen Wirtschaft nicht mit den Menschenrechten vereinbar sein sollten, könnte dies aufgrund von intransparenten Praktiken nicht erkannt und erst recht nicht sanktioniert werden.

Was ist zu tun, damit Algorithmen die Geschlechter unterscheiden und adäquat zuordnen können? Welche Herausforderungen sind dabei zu lösen?

Wir benötigen eine soziotechnische Perspektive auf die Digitalisierung hinsichtlich Gestaltung, Nutzung und Zugang. Automatisierte Prozesse oder algorithmische Systeme sind nie „neutral“, sie besitzen einen gesellschaftlichen Kontext. Die digitale Transformation ist also ein gesellschaftlicher Prozess, der nicht rein technikzentriert und/oder rein marktorientiert angegangen werden kann, ohne in Widerspruch zum Grundgesetz zu geraten.

Ist Gendergerechtigkeit vorrangig eine öffentliche Aufgabe oder auch privatwirtschaftlich relevant? Welche rechtlichen Vorgaben sollten diesbezügliche Algorithmen bekommen?

Öffentliche Stellen haben diesbezüglich bereits vieles in die Wege geleitet, wie beispielsweise die Umsetzungsstrategie „Digitalisierung gestalten“ und das Gutachten zum Dritten Gleichstellungsbericht der Bundesregierung zeigen. Als Richtschnur des

Handelns dient dabei Artikel 3 Abs. 2 des Grundgesetzes. Um die Gestaltung des digitalen Transformationsprozesses mit dem Verfassungsauftrag in Einklang zu bringen, ist die Umsetzung durch privatwirtschaftliche Akteur*innen zentral. Die Gesellschaft für Informatik gibt in ihrem Ethischen Kompass eine Orientierung.

DEBATTENBEITRAG

03.05.2023 | INTERVIEW

DISKRIMINIERUNGEN WIRKSAM BEGEGNEN

Geschlechter müssen für Datenerfassung und -verarbeitung sichtbar sein



Prof. Dr. Nicola Marsden, Inhaberin der Forschungsprofessur für Sozioinformatik an der Hochschule Heilbronn [Quelle: Monika Pröbster]

"Richtig gestaltet, erweitern KI-Systeme unsere Möglichkeiten, eine diskriminierungsfreie und geschlechtergerechtere Welt zu schaffen", sagt Prof. Dr. Nicola Marsden, Inhaberin der Forschungsprofessur für Sozioinformatik an der Hochschule Heilbronn und stellvertretende Vorsitzende des Kompetenzzentrums Technik-Diversity-Chancengleichheit. Und sie erklärt, wie das idealerweise umgesetzt werden kann.

Inwieweit ist eine gendergerechte KI, sind intelligent programmierte und gendersensitive Algorithmen für Sie ein praktisch relevantes Problem?

Das ein absolut praktisch relevantes Problem, da KI und Algorithmen in immer mehr

Bereichen unseres Lebens zum Einsatz kommen. In unserer Gesellschaft geschieht Geschlechterdiskriminierung meist unbewusst: Wir tragen tradierte Geschlechtervorstellungen in uns, die im Alltag wirksam werden, ohne dass wir dies intendieren. Das sieht man zum Beispiel daran, dass Gründerinnen eher nach den Risiken ihres Vorhabens gefragt werden, Gründer hingegen eher nach den Potenzialen. Oder wenn Männern technische Kompetenz qua Geschlecht zugesprochen wird, während Frauen sie zuerst und dann immer wieder aufs Neue beweisen müssen. Wenn algorithmische Systeme nun aus den Daten der Vergangenheit lernen, dann besteht die Gefahr, dass wir genau diese Verzerrungen perpetuieren, was sich beim Einsatz von KI im Recruitmentprozess auch schon gezeigt hat. Dabei haben wir die Chance, in allen Phasen der Gestaltung – der Datenerhebung, der Modellbildung, der Nutzenden-Interaktion – gezielt darauf hinzuwirken, dass die menschlichen Biases eben nicht zum Tragen kommen, sondern wir tatsächlich faire und gendergerechte KI schaffen.

In welchen Bereichen sehen Sie eine besondere Dringlichkeit für entsprechende Lösungen?

Besondere Dringlichkeit gibt es natürlich überall dort, wo das Leben und die Existenz von Menschen direkt betroffen sind.

Sie haben oben den Gender-Data-Gap in der Medizin genannt. Es gibt aber auch den Gender-Diagnose-Bias, der heute in der ärztlichen Praxis dazu führt, dass bestimmte Krankheiten bei Frauen unterschätzt oder weniger gut diagnostiziert werden. Wenn wir KI-Systeme nun einfach auf Basis der existierenden Daten lernen lassen, übernehmen sie den Bias. Schlimmer noch: Sie reproduzieren und verschärfen ihn. Faire KI-Lösungen können hier Leben retten.

Wenn Frauen auf Jobplattformen schlechter bewertet werden, wenn sie bei Einstellungen benachteiligt werden oder wenn Videokonferenzsysteme ihre Stimme so übertragen, dass sie weniger kompetent erscheinen, dann geht es zwar vordergründig nicht um Leben und Tod. Aber es geht um Grundrechte und die Frage, ob wir verhindern möchten, dass unsere Vorurteile automatisiert in Systemen perpetuiert wer-

den.

Was ist zu tun, damit Algorithmen die Geschlechter unterscheiden und adäquat zuordnen können? Welche Herausforderungen sind dabei zu lösen?

Die Betrachtung nach Geschlecht ist wichtig, um Diskriminierungen wirksam zu begegnen. Auch und gerade wenn wir Gleichstellung wollen, dürfen wir Geschlecht nicht unsichtbar machen. Wir müssen sicherstellen, dass bei der Datenerfassung und -verarbeitung eine ausreichende Anzahl von Datenpunkten für alle Geschlechter vorhanden ist. Hierbei sollten auch Aspekte wie Alter, ethnischer Hintergrund und sozio-ökonomische Unterschiede berücksichtigt werden.

Das heißt also paradoxerweise: Wenn wir wollen, dass Geschlecht keine Rolle mehr spielt, müssen wir ihm besondere Aufmerksamkeit widmen. Nur dann können wir existierende Ungerechtigkeiten korrigieren. Es funktioniert nicht, Geschlecht einfach wegzulassen – dann kommt es über sogenannte Proxies wieder in die KI-Systeme hinein. Dafür wurden zum Beispiel Kreditvergabealgorithmen kritisiert. Sie hatten Geschlecht explizit nicht beachtet, was dann aber über mit Geschlecht korrelierenden Faktoren zu einer geringeren Kreditvergabe an Frauen geführt hat.

Ist Gendergerechtigkeit vorrangig eine öffentliche Aufgabe oder auch privatwirtschaftlich relevant? Welche rechtlichen Vorgaben sollten diesbezügliche Algorithmen bekommen?

Sowohl als auch. Die meisten Unternehmen möchten Geschlechtergerechtigkeit aktiv umsetzen und dazu beitragen, Diskriminierung zu minimieren und eine inklusivere Gesellschaft zu fördern. Sie tun dies aus sozialer Verantwortung und um ihr Geschäft zu verbessern.

Aktuell gibt es im Rahmen der Vorbereitung des AI-Acts, der europäischen Regelungen für Transparenz und Verantwortlichkeit von KI-Systemen, eine Diskussion über die Aufwände, die solche Regelungen mit sich bringen. Aber: Niemand von uns möchte, dass über unser Leben Entscheidungen getroffen werden, die möglicherweise

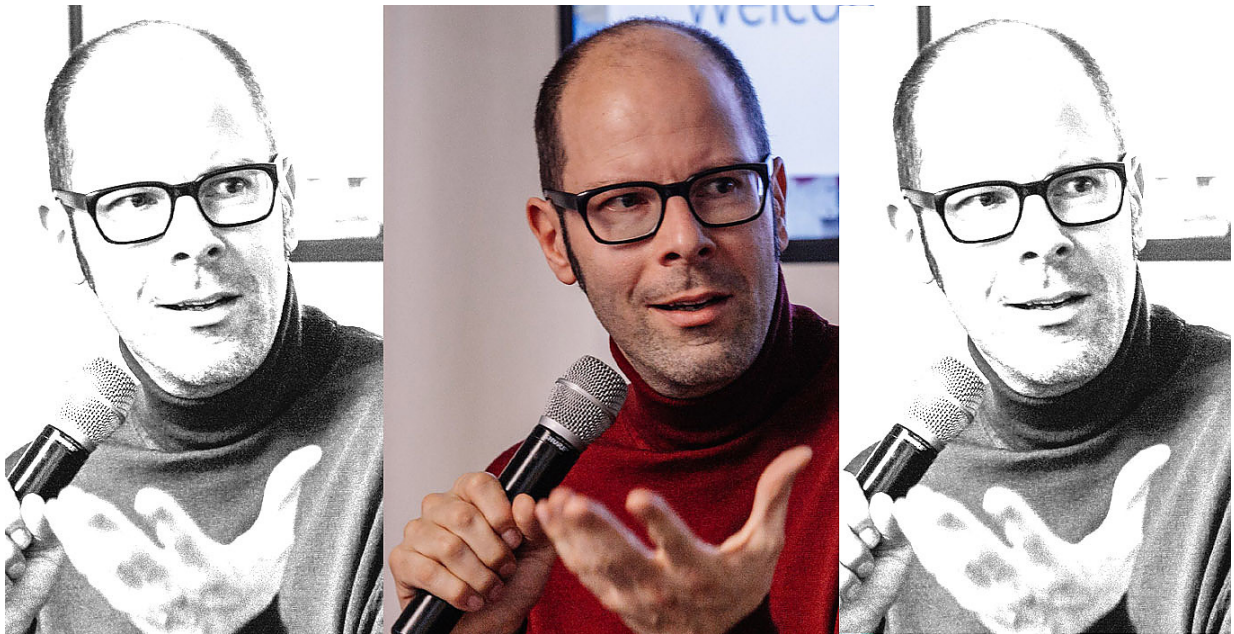
unfair sind und die nicht erklärbar sind. Deshalb gilt es, unaufgeregt zu überlegen, wie wir Lösungen schaffen, die unserem Verständnis von Transparenz, Verantwortlichkeit und Datenschutz gerecht werden. Denn, richtig gestaltet, erweitern KI-Systeme unsere Möglichkeiten, eine diskriminierungsfreie und geschlechtergerechtere Welt zu schaffen – dieses Potenzial sollten wir nutzen.

DEBATTENBEITRAG

05.05.2023 | INTERVIEW

WO TECHNOLOGINNEN UND TECHNOLOGEN IHRE GRENZEN KENNEN

Wie man Künstliche Intelligenz geschlechtergerecht machen kann



Aljoscha Burchardt [Quelle: Quelle]

Um Künstliche Intelligenz von vornherein gendgerecht aufzustellen, muss man zunächst einmal "ein Sensorium dafür haben, dass überhaupt eine Benachteiligung stattfinden kann", sagt Dr. Aljoscha Burchardt, Principal Researcher und stellvertretender Standortsprecher am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) in Berlin.

Inwieweit ist eine gendgerechte KI, sind intelligent programmierte und gendersensitive Algorithmen für Sie ein praktisch relevantes Problem?

Vorab erlauben Sie mir vielleicht einen Kommentar zur Frage selber. Es wird oft in der

Berichterstattung nicht differenziert, an welcher Stelle im komplexen KI-Entwicklungsprozess etwas schief gelaufen ist, bzw. an welcher Stelle Maßnahmen greifen sollten, damit KI-Systeme kein diskriminierendes bzw. benachteiligendes Verhalten zeigen. Das reicht von der Auswahl und Programmierung von Algorithmen über die Auswahl und das Training mit Daten bis hin zur Gestaltung der Mensch-Maschine-Schnittstelle und schließlich der Evaluation und Einführung von Technologie. Daher bin ich ein wenig über die Formulierung „intelligent programmierte Algorithmen“ gestolpert. Die Herausforderung ist vielschichtig.

Auf der Ebene individueller Forschungsprojekte spielen diese Fragen regelmäßig bei uns in ganz unterschiedlichen Kontexten eine Rolle. Dann holen wir typischerweise Experten mit ins Boot wie Ethikerinnen oder Patientenvertreter, da kennen wir als Technologinnen und Technologen unsere Grenzen.

Unsere zentrale Arbeitsgruppe Diversity & Gender Equality hat sich als ein Ziel gesetzt, genau die genannten Aspekte in der Forschung systematisch zu beleuchten, das Bewusstsein bei allen Mitarbeiterinnen und Mitarbeitern wach zu halten und mittelfristig auch aktiv an der Verbesserung des Forschungsstandes in diesen Fragen beizutragen.

In welchen Bereichen sehen Sie eine besondere Dringlichkeit für entsprechende Lösungen?

Als Technologie möchte ich mir gar nicht anmaßen, hier eine Priorisierung vorzunehmen oder bestimmte Beispiele zu nennen und dabei sicherlich andere zu vergessen. Es ist grundsätzlich besonders wichtig, wenn wir es mit vulnerablen Gruppen zu tun haben, die sich möglicher Weise gar nicht wehren können, wenn sie ungerecht behandelt werden. Und es kommt natürlich auch auf die „Fallhöhe“ an: Geht es um existentielle Fragen? Wobei etwa bei Medizinprodukten oder Dienstleistungen der Daseinsvorsorge ja auch die gesetzlichen Grundlangen wie das AGG gelten.

Was ist denn konkret bei der KI-Entwicklung zu tun, damit kein Geschlecht von

Algorithmen und Modellen benachteiligt wird? Welche Herausforderungen sind dabei zu lösen?

Das ist wirklich – und der Ausdruck ist ja auf seine Art sehr passend – die Gretchenfrage! Erst einmal muss man ein Sensorium dafür haben, dass überhaupt eine Benachteiligung stattfinden kann. Idealerweise ist man „Gender-neutral by Design“. Ein wichtiger Hebel sind hierbei die Trainingsdaten. Da stellt sich dann häufig die Frage: Wenn meine Daten nicht ausgewogen sind, dann habe ich erstmal die Wahl: Will ich das insgesamt bessere System, das dann vielleicht für Männer oder Frauen im Vergleich besser funktioniert oder nutze ich nur einen ausgewogenen Teil der Daten und gleiche damit das System dann „nach unten“ an. Dann gibt es Möglichkeiten, synthetische Daten zu erzeugen oder vielleicht, sofern sie irrelevant sind, Geschlechtsmerkmale ganz aus den Daten zu entfernen. Das ist aber auch nicht trivial, beispielsweise geben Elternzeiten in Lebensläufen typischer Weise Hinweise auf ein bestimmtes Geschlecht. Das sind dann sogenannte Proxy-Merkmale. Hier beginnt dann der spannende Teil der Forschung.

Ist Gendergerechtigkeit vorrangig eine öffentliche Aufgabe oder auch privatwirtschaftlich relevant? Welche rechtlichen Vorgaben sollten diesbezügliche Algorithmen bekommen?

Wie ich bereits sagte, sind bestimmte Bereiche ja ohnehin schon reguliert. In der Enquete-Kommission „KI“ des Deutschen Bundestages haben wir seinerzeit viel über diese Fragen diskutiert. Insgesamt spielen so viele Aspekte hinein: Wer bestimmt, was Gleichheit ist? Wir kennen aus den Sozialen Medien den schmalen Grat zwischen Schutz und Zensur. Wer kontrolliert die Privatwirtschaft? Staatliche Akteure, halbstaatliche? Die Antwort fällt in Deutschland wahrscheinlich anders aus als in China. Die komplexen Antworten auf alle diese Fragen müssen wir in einem gesamtgesellschaftlichen Diskurs aushandeln.

DEBATTENBEITRAG

15.05.2023 | INTERVIEW

WIE KI-ANWENDUNGEN STEREOTYPEN REPRODUZIEREN

Und warum auch Algorithmen gendersensitiv sein müssen



Nadia Fischer, CEO und Co-founder Witty Works [Quelle: Witty Works]

"Sprache ist nicht neutral. Wörter, die wir nutzen, um unsere Umwelt zu beschreiben, bestimmen, wie wir diese Welt sehen und empfinden", erklärt Nadia Fischer den Ausgangspunkt ihrer Unternehmung: Sie ist CEO & Co-founder des KI-basierten digitalen Schreibassistenten Witty, der in Wirtschaft und Gesellschaft für mehr Inklusivität sorgen soll.

Inwieweit ist eine KI, die nicht gendersensitiv sind, ein praktisch relevantes Problem?

Ich untersuche das Problem vorrangig in Bezug auf "Large Language Models" (LLM). Also grosse Sprachmodelle, mit denen heute viele KI betrieben werden, wie z.B.

ChatGPT von OpenAI / Microsoft oder Bard von Google. Tatsächlich wurde in den verschiedensten Studien* wie auch in einfachen Experimenten festgestellt, dass diese Modelle 'voreingenommen' sind. D.h. sie behandeln die Geschlechter nicht gleich, sondern reproduzieren Stereotypen. Sogar in Sprachen, die linguistisch neutral sind, wie z.B. englisch, werden solche Stereotypen wiederholt.

Zwei Beispiele:

1. In deutsch: Wenn wir in Deepl eine Übersetzung von 'kindergarden teacher' auf deutsch anfragen, so wird diese automatisch als weibliche Person übersetzt, obwohl die englische Version in keiner Weise darauf hin deutet.

2. In englisch: Wenn man ChatGPT Fragen stellt in Bezug auf Arzt (in englisch: doctor; genderneutraler Begriff) und Pflegeverantwortliche (in englisch: nurse, genderneutraler Begriff), so reproduziert ChatGPT den Stereotyp, dass ein 'doctor' ein Mann sein muss und eine 'nurse' eine Frau. Wenn man ChatGPT darauf hinweist, dass eine 'nurse' ein Mann sein könnte, dann widerspricht hier ChatGPT und meint, es handle sich wahrscheinlich um einen Logik- oder Grammatik-Fehler.

LLMs lernen von menschlichem, auf dem Internet publizierten und in der Vergangenheit geschriebenen Inhalt. Diese Daten werden dem Algorithmus gefüttert. Dieser lernt dann aufgrund der Daten selbst, welches die Regeln des Inhaltes sind und wendet diese bei der Produktion von neuen Texten an. Je nach KI geht es darum, einen Text zu übersetzen (deepL) oder eben selbst Text zu schreiben (chatGPT). Es gibt noch viele weitere Anwendungsbereiche, die zurzeit einen grossen Boom erleben. In gewissen Fällen werden auch sogenannte 'Supervisor' dazwischengeschaltet, die dem Algorithmus anfangs helfen, Daten zu verstehen. Solche Supervisor bestimmen zum Beispiel, ob ein Begriff eine rassistische Konnotation hat. Denn ein LLM hat kein ethisches Verständnis.

Die Frage ist, wieso ist es ein Problem, wenn solche Sprach-KI Stereotypen wie-

der gibt?

Man könnte ja einfach sagen, dass das eben Stereotypen oder sogar Tatsachen sind, die es einfach gibt in unserer Gesellschaft. Z.B. eine Mehrzahl der Krankenpfleger sind nun mal Frauen. Diese Argumentation scheint mir viel zu gefährlich, denn sie nimmt in Kauf, dass wir uns als Gesellschaft nicht vorwärts bewegen wollen; dass wir uns nicht entwickeln können. Sprache hat sich immer angepasst und gemäss der gesellschaftlichen Entwicklung verändert. Sprache ist nicht nur ein Symptom unserer Gesellschaft, sondern auch Vorreiterin. Wenn neue Erfindungen gemacht werden oder neue Phänomene entstehen, dann entsteht dafür ein Wort, meist zuerst durch eine kleine Gruppe. Diese trägt das neue Wort in die Welt hinaus und - sofern die Erfindung einigermaßen erfolgreich ist - wird es in den allgemeinen Wortschatz übernommen. Hinter diesem Wort verbirgt sich eine neue Vorstellung, z.B. die Lösung eines Problems oder ein neuer Trend. Nun verstehen alle das neue Konzept und die Gesellschaft hat dabei einen Fortschritt gemacht - nicht nur in der Sprache, sondern auch im Verständnis des Problems oder des Trends. Um das etwas klarer zu machen: In der Schweiz sind 60% der Medizin-Absolvierenden Frauen. Dass also die KI besagt, dass ein 'doctor' ein Mann sein muss, ist also grundfalsch. Das LLM kann den Trend nicht erkennen, weil sie auf alten Daten aufbaut und eine falsche Regel gelernt hat.

Nun liegt das LLM nicht nur falsch, sondern mit diesen falschen Annahmen kreiert es auch eine Realität, die wir nicht mehr haben wollen. Denn Sprache ist nicht neutral. Wörter, die wir nutzen, um unsere Umwelt zu beschreiben, bestimmt, wie wir diese Welt sehen und empfinden. Sprechen wir also nur von Krankenschwestern, dann sehen wir vor dem inneren Auge auch nur Frauen. Formulieren wir dies jedoch um zu Krankenpfleger*innen, so stellen wir plötzlich fest, dass wir uns auch andere Geschlechter in diesem Beruf vorstellen können. Damit ermöglichen nicht nur wir uns selbst eine aufgeschlosseneren Sichtweise, sondern öffnen auch die Türen für Menschen, die diesen Beruf wählen wollen, aber keine Frauen sind. Denn damit fallen die Stereotypen weg und die Stigmata, die dem Beruf anfallen.

Dass ein LLM alte Stereotypen übernimmt, ist problematisch, weil die von LLM produ-

zierten Texte meist ohne weitere Überprüfung von den Nutzenden übernommen werden. So verbreiten sich Stereotypen weiterhin, obwohl wir sie als Gesellschaft loshaben wollen. Da auch Sprach-KI einen echten Boom erlebt, können wir annehmen, dass ein Vielfaches an Texten ins Internet zurückgespielt wird, die von KI geschaffen wurde. Damit amplifiziert sich sogar der rückwärtsgewandte Trend.

Und das ist ein Problem: Für uns als Gesellschaft und für viele soziale Gruppen. Wenn voreingenommene Sprach-KI immer öfters ungeprüft genutzt wird, dann werden sich Stereotypen in der Sprache wieder stärker durchsetzen. Durch die Wirkung der Sprache auf die Gesellschaft bedeutet dies, dass wir selbst wieder in mehr Stereotypen beginnen zu denken. Und das ist eine Rückwärtsbewegung, die Ungleichstellung zwischen den Geschlechtern verstärkt und das Geschlechterverständnis der neuen Generationen nicht berücksichtigt.

In welchen Bereichen sehen Sie ein besonderes Risiko der LLM?

Texte, die mit LLM erschaffen wurden, können für Unternehmen mehrere Nachteile haben und sogar Risiken bergen, wenn sie unüberprüft übernommen werden. Abgesehen vom Datenschutz, sind dies die folgenden:

Nutzen wichtige Persönlichkeiten wie Geschäftsleitungsmitglieder oder Politiker*innen solche Texte, schleichen sich Begriffe oder Ausdrücke ein, die Diskriminierung ausdrücken oder einen Stereotypen wiederholen. Verschiedene Communities und die Generation Z hat ein scharfes Auge auf solche voreingenommenen Aussagen. Inzwischen führen sie in den USA sofort zu einem PR-Fiasko. Die Vertreter*innen von Politik und Wirtschaft müssen zurücktreten oder werden entlassen. Teilweise führte es sogar zu Rechtsverfahren. Dies haben wir bis anhin noch nicht gesehen in Europa. Aber es gibt die ersten Anzeichen dafür, dass Partner Verträge kündigen, wenn solche diskriminierende Aussagen durch GL-Mitglieder gemacht werden, weil ihre Marke damit in Verbindung gebracht wird. D.h. die von der KI vorgegebenen Texte stellen für Unternehmen ein Risiko dar - für Führungskräfte wie auch für die Marke selbst.

In Zeiten des Fachkräftemangels ist es wichtig, dass alle potentiellen Talente angesprochen werden. LLM-produzierte Texte verwenden jedoch einerseits, wie oben gesehen, Stereotypen in Bezug auf die Titel; andererseits produziert Chat GPT Stellenanzeigen, die 'traditionell' geschrieben sind. D.h. mit Ausdrücken, die Leistung und Wettbewerb zum Ausdruck bringen. Aus Studien der TU München und Harvard wissen wir jedoch, dass weibliche Talente nicht gut auf solche Wortwahl ansprechen und sich dann eher nicht bewerben. Werden also LLM-produzierte Texte genutzt für die Stellenanzeigen, Employer Branding Texte oder für das Active Sourcing, dann werden die Bewerbungen nicht so vielfältig ausfallen, wie sie eigentlich könnten. Durch die Sprache werden potentielle Talente abgeschreckt.

Im Marketing muss heute ein sehr breiter Fächer von Zielgruppen angegangen werden. Menschen in Europa werden immer älter und die Globalisierung führt zu einer wachsenden Diversität. Aufgrund der letzten Volkszählung in der USA wird prognostiziert, dass die "Weissen" in den USA in 2045 zur Minderheit werden. Für Unternehmen bedeutet dies auch diesseits des Atlantik, dass sie all diese Zielgruppen in ihrem Marketing ansprechen können müssen, um wettbewerbsfähig zu sein. ChatGPT wird heute im Marketing zunehmend verwendet. Eine kürzlich durchgeführte, nicht repräsentative Studie von FreelancerMap im deutschsprachigen Raum hat ergeben, dass 61 % der 854 befragten Freiberufler*innen regelmäßig KI bei ihrer Arbeit einsetzen, und 40 % von ihnen verwenden Sprach-KI, um bei der Erstellung von Texten zu helfen. Wenn nun diese Texte unüberprüft übernommen werden und damit auch die Stereotypen und unbewussten Diskriminierungen bedeutet dies für Unternehmen, dass sie mögliche Kund*innen abschrecken. Sie verlieren also damit Marktpotenzial.

Ich empfehle darum jeweils eine Überprüfung der Resultate von Sprach-KI mit dafür gemachten Tools, die Voreingenommenheiten und Stereotypen in den LLM-Texten erkennen können. Diese Tools werden 'AI Whisperer' genannt. Sie flüstern den LLM-Texten gender-neutralere und unvoreingenommene Lösungen ein. Witty (<https://witty.works>) ist ein Beispiel eines solchen Tools.

Was ist zu tun, damit Algorithmen die Geschlechter unterscheiden und adäquat zuordnen können? Welche Herausforderungen sind dabei zu lösen?

Die Idee der LLM ist, dass der Algorithmus selbst von einer grossen Datenmenge lernen kann. Um also in seinem Umgang mit Geschlechtern besser zu werden, sind drei Faktoren wichtig:

1. Die Trainings-Daten müssen schon bereinigt sein. D.h. es muss zuerst überprüft werden, ob die Daten selbst schon Voreingenommenheiten enthalten oder Stereotypen beinhalten. Dies ist noch einfach, wenn man sich eine kleine Datenmenge vorstellt und es nur um Geschlechter geht. Es wird schwierig, wenn mit sehr grossen Datenmengen gearbeitet wird ('aller Content des Internet') und nicht nur Voreingenommenheit in Bezug auf Geschlecht, sondern auch in Bezug auf z.B. sozio-ökonomische Schichten ausgeräumt werden sollen.
2. Die Modelle werden von Teams erarbeitet. Sind diese Teams homogen aufgestellt, dann werden ihnen Voreingenommenheit in ihren Entscheidungen weniger auffallen. Dies wiederum bedeutet, dass die Modelle auf Unvoreingenommenheiten nicht sensibilisiert sind. Um dies zu verhindern, sollten die Teams selbst divers aufgestellt sein.
3. Die Supervisor selbst sollten auch divers sein, damit sie möglichst viele Diversitätsdimensionen abdecken können und somit die Daten dementsprechend unter die Lupe nehmen.

Ist Gendergerechtigkeit vorrangig eine öffentliche Aufgabe oder auch privatwirtschaftlich relevant? Welche rechtlichen Vorgaben sollten diesbezügliche Algorithmen bekommen?

Gendergerechtigkeit ist privatwirtschaftlich relevant. Studien von Boston Consulting Group, McKinsey, Deloitte wie auch der ILO bezeugen, dass divers aufgestellte Unternehmen innovativer und langfristig erfolgreicher sind, da die Entscheidungen von heterogenen Teams gefällt werden. Will ein Unternehmen nachhaltig erfolgreich sein, ist Diversität im Team ein Business-Imperativ. Zudem bedeutet Gendergerech-

tigkeit im Unternehmen mehr Fairness für alle. Denn wenn Frauen fair behandelt werden, hat dies einen positiven Einfluss auf andere unterrepräsentierte Gruppen. Für das Unternehmen selbst ist Gendergerechtigkeit auch kurzfristig von Interesse: Frauen fühlen sich dadurch im Unternehmen angenommen. Sie sind produktiver, bleiben länger beim Unternehmen und sprechen gegen aussen gut über das Unternehmen. Alles Folgen, die sich positiv auf das Geschäftsergebnis auswirken.

Es ist daher wichtig, dass es mehr rechtliche Vorgaben gibt für Sprach-KI. Zurzeit befinden wir uns in einem fast rechtlosen Raum. Folgende Bestimmungen wären wichtig:

- Die KI darf Menschen weder offensichtlich noch unbewusst diskriminieren aufgrund ihres Geschlechts, Herkunft, sexuelle Orientierung, Alter, usw. So wie es in den meisten Verfassungen hinterlegt ist.
- Das Unternehmen muss offenlegen, welche Daten für das Training des Algorithmus genutzt werden.
- Supervisor sowie Teams müssen einen gewissen Grad an Diversität ausweisen.
- Zu grosser Einfluss durch einzelne Unternehmen sollte verhindert werden, da sonst ihre Modelle zu vorherrschend werden.

Diese Vorschläge sind nicht abschliessend gemeint.

* siehe eine Zusammenfassung hier: <https://www.witty.works/en/blog/is-chatgpt-able-to-generate-inclusive-language>

DEBATTENBEITRAG

20.06.2023 | STATEMENT

ETHIK IST NICHT DIE FAHRRADBREMSE AM INTERKONTINENTALFLUGZEUG

Warum eine ethische Perspektive notwendig ist, um
Diskriminierungsmechanismen offenzulegen



Prof. Dr. Regina Ammicht-Quinn, Universität Tübingen [Quelle: Margret Garbrecht, Tübingen]

"Wir brauchen mehr Frauen und Minderheiten im KI-Bereich, aber wir brauchen genauso ein größeres Bewusstsein für Diskriminierungsvorgänge bei allen, die in diesem Bereich arbeiten und bei allen, die die Wirtschaft dieser Bereiche beeinflussen", sagt Prof. Dr. Regina Ammicht Quinn. Sie ist unter anderem Sprecherin und Leiterin des Bereichs Gesellschaft, Kultur und technischer Wandel des Wissenschaftlichen Rats des Internationalen Zentrums für Ethik in den Wissenschaften (IZEW) der Universität Tübingen.

Von den 1940er bis zu den 1960er Jahren war Programmieren Frauenarbeit. Frauen

galten als besonders qualifizierte Programmiererinnen (schließlich, so wurde argumentiert, waren sie auch in der Lage, Kochrezepte und Strickmuster zu entwerfen). Die Arbeit hatte noch nicht den Status, den sie heute hat, der Begriff „Software“ war noch nicht erfunden und „coding“ galt als sekundäre Aufgabe, während Ruhm und Ehre den Herstellern der Maschinen zukam. Dass es heute eklatant wenige Frauen in diesem Bereich gibt, hat also wenig mit der DNA oder anscheinend geschlechtsspezifischen Begabungen zu tun, sondern mit einer Sozialgeschichte, die im Einzelnen noch nachgezeichnet werden muss.

Vergangenheit, Gegenwart ... und Zukunft?

In einem Land wie Deutschland, das zu den „KI-Nationen“ gehört, hat eine solche Schieflage erhebliche Konsequenzen: die asymmetrische ökonomische Partizipation von Frauen, der dringende Bedarf an Fachkräften, der nicht gedeckt werden kann und schließlich auch ein Qualitätsproblem. Wenn die Expertise und Erfahrung sehr vieler Menschen keinen Eingang in die Technologieentwicklung und -Implementierung finden, diese Technologien aber Leben und Gesellschaft entscheidend prägen, dann haben wir ein Problem.

Fragen der Geschlechtergerechtigkeit sind intersektionale Fragen. Das bedeutet, dass es nicht allein um „Frauen“ geht. Unterschiedliche Diskriminierungsmechanismen müssen zusammen gedacht werden: Die Benachteiligung aufgrund des Geschlechts ist verknüpft mit der Benachteiligung aufgrund des sozioökonomischen Status und der Benachteiligung aufgrund von Herkunft und Hautfarbe.

Und was hat die Ethik damit zu tun?

Ethik ist nicht, wie Ulrich Beck es formulierte, die „Fahrradbremse am Interkontinentalflugzeug“. Ethik bremst nicht den Fortschritt, sondern fragt vielmehr, was Fortschritt ist und wie er gestaltet werden sollte. Aus einer ethischen Perspektive geht es darum, dass die Prinzipien, die grundlegend sind für eine demokratische Gesellschaft, übersetzt werden müssen für eine digitale demokratische Gesellschaft. Dazu braucht es neue Prozesse und Strukturen, etwa Verantwortung, Zurechenbarkeit, und gute

Regulierung. Und es braucht das Wissen darum, wer von einer Technologie profitiert, wer einem Einsatz widersprechen kann und in welcher Weise neue Abhängigkeiten geschaffen werden können. Momentan zeigen sich die Probleme insbesondere als strukturelle Probleme, als Probleme der Repräsentation und als Gerechtigkeitsprobleme.

Probleme, Lösungen und die Problematik von Lösungen

Das strukturelle Problem der geringen Zahl an Frauen und Minderheiten im KI-Feld wird häufig als „Pipeline“-Problem behandelt. Man versucht, möglichst früh möglichst viele Mädchen und Frauen (und Minderheiten) in eine Pipeline hineinzustopfen. Ermutigung und Bildungsprogramme sind keineswegs falsch. Leider sind die Ergebnisse mittelmäßig. Die Pipeline ist undicht, und Frauen gehen unterwegs verloren. Das scheint deren Schuld zu sein: Sie können es nicht, sie wollen nicht, oder sie passen nicht. Ein so wahrgenommenes „Pipeline“-Problem vermeidet es nachzusehen, wie es im Inneren der „Pipeline“ aussieht und welche alternativen oder ungewöhnlicheren Wege es für die Karrierewege von Frauen und Minderheiten gibt und geben könnte. Wenn ein hoher Anteil an Frauen nach 10-15 Jahren das Arbeitsfeld KI wieder verlässt, dann müssen nicht nur Frauen und Minderheiten sich verändern, sondern genauso der existierende Mainstream.

Das Problem der mangelnden Repräsentation zeigt sich beispielweise in medizinischen Daten und endet noch nicht bei Daten für Stadtplanung oder Verkehrssicherheit: Daten und Datenanalysen liegen Entscheidungen zugrunde, und Frauen und andere, die nicht zum „Mainstream“ gehören, sind unterrepräsentiert oder missrepräsentiert. Mehr Daten, etwa von indigenen Frauen für Gesundheit und Bildung, sind entscheidend; denn nur das, was gezählt werden kann, zählt letztlich auch für Forschung, für die Zuteilung von Mitteln, für Unterstützung oder die Anerkennung von Rechten. Zugleich aber entsteht hier ein „paradox of exposure“ – ein Paradox der Bloßstellung: Gerade diejenigen, die besonders davon profitieren könnten, wenn sie „gezählt“ werden, sind häufig in besonderer Gefahr durch eben diese Datenerfassung und Klassifizierung benachteiligt zu werden. Dazu gehören Migrant:innen, Menschen,

die als „illegal“ gelten oder auch trans*Menschen, die konstant mit Angriffen zu rechnen haben.

Das Problem der Gerechtigkeit, im technischen Kontext als „bias“, also „Verzerrung“, benannt, ist Technikentwickler:innen bewusst. Technische Lösungen (DADM, FAccTML und andere) werden beständig weiterentwickelt, sodass mit „Fair AI“ ein neues Forschungsfeld entstanden ist. Hier ist eine Definition von „bias“ nötig, die leicht operationalisiert und bis zu einem gewissen Grad korrigiert werden kann, beispielsweise dort, wo die Fehlerraten für eine einzelne Variable ungleich verteilt sind.

Es gibt aber Benachteiligungen, die komplexer sind und keine besondere Feindseligkeit oder Bevorzugung einzelner Menschen enthalten. Hier werden Menschen bestimmten Gruppen zugeordnet, und wahrscheinlichkeitstheoretische Einschätzungen von Gruppencharakteristika werden genutzt, um individuelle Charakteristika zu evaluieren. Ob die Gruppe „Frau“ oder „Trans*“ oder „Schwarz“ oder „Postleitzahl“ heißt – hier werden Gerechtigkeitsprobleme generiert. Sichtbar wird dies nicht nur bei Predictive Policing, bei der Vergabe von Krediten, Studienplätzen (usw.), sondern auch bei Systemen, die die Vernachlässigung oder den Missbrauch von Kindern verhindern sollen. Die Daten, die hier zugrunde liegen, sind beispielsweise Anträge auf Sozialhilfe und auf andere Unterstützung, kontinuierliche externe Überprüfungen und Polizeidaten. Arme Kinder tauchen wesentlich häufiger in den Datensätzen auf als andere. Damit sind (teil)automatisierte Systeme in Gefahr, „schlechte“ Eltern mit „armen“ Eltern zu verwechseln.

Wir brauchen also mehr Frauen und Minderheiten im KI-Bereich, aber wir brauchen genauso ein größeres Bewusstsein für Diskriminierungsvorgänge bei allen, die in diesem Bereich arbeiten und bei allen, die die Wirtschaft dieser Bereiche beeinflussen. Damit müsste KI ein interdisziplinäres Feld werden, das auch die Geistes- und Gesellschaftswissenschaften mit einbezieht. Die Entwicklung von Technologien, die innerhalb von komplexen sozialen Realitäten agieren sollen, erfordert ein kritisches Verständnis sozialer, rechtlicher und ethischer Kontexte und deren Interdependenz.

Digitale Hegemonien

Faktoren, die lange Zeit das Leben von Frauen bestimmt haben, und mittlerweile in den Geschlechterstudien diskutiert und kritisiert werden, finden sich auch in der Entwicklung von KI-Systemen wieder: unsichtbare Arbeit, Materialität (statt „Geist“) und kaum geschätztes Wissen. Hier können die KI-Analysen von feministischen Analysen lernen.

Eine der wenig wahrgenommenen Tatsachen ist die unsichtbare und unterbezahlte Arbeit, die KI-Systemen zugrunde liegen: monotone digitale Arbeit, labeling, und das „Säubern“ der Datensätze von schädlichen oder gewalttätigen Inhalten. Lilly Irani nennt dies: durch Menschen betriebene Automatisierung.

Materialität, Körperlichkeit und „Natur“ waren die Zuschreibungen für Frauen. KI gilt als immateriell und „sauber“. Aber es ist an der Zeit, zum einen den ökologischen Fußabdruck zu beziffern, zum anderen die Lieferketten offenzulegen, auch die Extraktion von seltenen Erden durch Kinderarbeit und Arbeit in Krisen- und Kriegsgebieten.

Was „Wissen“ bedeutet und was „Fakten“ sind, ist in westlichen Kontexten festgelegt worden. „Fake news“ sind ein aktuelles großes Problem – aber nicht das einzige Problem. Generative KI wie ChatGPT und die Nachfolgemodelle generieren aus vorhandenen Datensätzen überschaubares Wissen, das wiederum in die Wissensbasis eingeht. Aber welches und wessen Wissen ist das? Und welches Wissen wird ausgeschlossen? Die Übermacht eines in Form und Inhalt „westlichen“ Wissens kann ein globales System wie KI in ein koloniales Werkzeug verwandeln.

DEBATTENBEITRAG

23.08.2023 | STATEMENT

ÜBER DIE ILLUSION VON OBJEKTIVITÄT BEI DER NUTZUNG VON KI-SYSTEMEN

Wie KI die Realität richtig erkennen kann



Prof. Dr. Sabine T. Köszegi vom Institute of Management Science der Technischen Universität Wien [Quelle: Luiza Puiu]

Geschlechterstereotypen schlagen sich auch in algorithmischen Systemen nieder - was zur Perpetuierung von Stereotypen und Vorurteilen führt, ungerechtfertigter Diskriminierung und Fehlentscheidungen, erklärt Prof. Dr. Sabine T. Köszegi vom Institute of Management Science der Technischen Universität Wien.

Hinter dem Begriff Künstlicher Intelligenz verbergen sich in der Regel algorithmische Systeme, die mithilfe verschiedener Methoden des maschinellen Lernens große Datenmengen analysieren. Algorithmen sind demnach das Ergebnis einer Modellbildung, die sowohl die Formalisierung eines Problems als auch eine entsprechende

Zielsetzung beinhalten. Für die Effektivität von KI Systemen sind die Werte und Annahmen der Entwickler:innen von ebenso großer Bedeutung wie die dem zugrundeliegenden Daten und der Algorithmus, der aus ihnen lernt. Soziale Urteile darüber, wer oder was für eine Problemstellung relevant ist, wird von Systementwickler:innen durch Designentscheidungen und Operationalisierung von strategischen Zielen vorweggenommen. Wenn wir von Algorithmen sprechen, beziehen wir uns also genau genommen auf ein sozio-technisches Ensemble von technischen Anordnungen, in der die Beteiligung von Menschen aber verdeckt bleibt. Mit der Bezeichnung „Algorithmus“ wird verschleiert, dass kulturelle, gesellschaftliche und politische Werte lediglich in eine mathematische Sprache übersetzt sind und es entsteht eine Illusion von Objektivität bei der Nutzung von KI-Systemen, die nicht gerechtfertigt ist. Denn wir Menschen neigen dazu, vermeintlich objektiven KI-Systemen mehr zu vertrauen, als potenziell fehleranfälligen, mit Vorurteilen behafteten, emotionalen Menschen.

In algorithmischen Systemen können systematische Verzerrungen („Bias“) in jeder Phase ihres Lebenszyklus – von der Konzeptions- über die Entwicklungs-, Implementierungs-, Nutzungs- und Evaluierungsphase einfließen und damit zu Perpetuierung von Stereotypen und Vorurteilen, ungerechtfertigter Diskriminierung und Fehlentscheidungen führen. Das Berkeley Haas Center for Equity, Gender, and Leadership hat 133 Systeme aus den Jahren 1988 und 2021 branchenübergreifend analysiert und festgestellt, dass ein alarmierend hoher Anteil von 44,2 Prozent der Systeme einen geschlechtsspezifischen Bias aufweist. Etwa ein Viertel der Systeme weist sowohl geschlechts- als auch rassistische Vorurteile auf. In der Tat kommt der AI NOW Report (2018) zu dem Schluss, dass die Schäden und Verzerrungen in KI-Systemen außer Frage stehen: "Die Debatte ist entschieden, die Beweise sind über jeden Zweifel erhaben". Es stellt sich heraus, dass algorithmische Entscheidungen aufgrund unvollständiger oder unvollständiger Daten, unzureichender Modellierung und problematischer Zielsetzungen mit schwerwiegenden Problemen verbunden sind und im Grunde mit schwerwiegenden Folgen für den einzelnen und die Gesellschaft scheitern.

Der UNESCO Bericht 2018 „I'd blush if I could“ greift insbesondere die Bildungsunterschiede zwischen Frauen und Männer in den MINT und ICT-Bereichen auf und zeigt, wie sich Geschlechterstereotypen in segregierten Bildungsentscheidungen niederschlagen und dazu führen, dass Frauen massiv in der MINT und ICT Bildung als auch in den technischen Professionen unterrepräsentiert sind. Dies verschlechtert nicht nur gesellschaftliche Teilhabechancen von Frauen, sondern führt auch dazu, dass KI-Technologien hauptsächlich von Männern für Männer entwickelt werden – mit den Konsequenzen, die wir aus den oben angeführten Berichten kennen.

Um diesen Teufelskreis zu durchbrechen bedarf es daher massiver Anstrengungen – auch im Sinne der Qualität und Effektivität von KI Technologien – bei Entwickler:innen Teams einen Fokus auf Gendergerechtigkeit, Diversität und Inklusion zu setzen.

Die Potenziale von KI-Technologie können nur dann gehoben werden, wenn sie gendersensibel und inklusiv, die Menschenrechte achtend gestaltet ist. Das ist ausschließlich eine Frage unserer Werte und keine Frage der technischen Machbarkeit.

IMPRESSUM

Herausgeber

Barthel Marquardt GbR

Merseburger Straße 200

04178 Leipzig

Tel: 0341 24 66 43 72

E-Mail: marquardt@meinungsbarometer.info

www.meinungsbarometer.info

V.i.S.d.P.

Dipl.-Journ. Nikola Marquardt

Idee, Konzept, Projektleitung

Dipl.-Journ. Thomas Barthel

Redaktion

Barthel Marquardt GbR

Diese Dokumentation darf nicht - auch nicht in Auszügen - ohne schriftliche Erlaubnis der Redaktion vervielfältigt und verbreitet werden. Die Dokumentation wurde mit größtmöglicher Sorgfalt zusammengestellt. Trotzdem können wir für die enthaltenen Informationen keine Garantie übernehmen. Die Redaktion schließt jegliche Haftung für Schäden aus, die aus der Nutzung von Informationen dieser Dokumentation herrühren.