

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

Hi, ich bin's Victoria. Ab Montag gibt es wieder neue Folgen von 11KM der Tagesschau-Podcast. Und damit heute meine letzte Empfehlung aus den bisherigen Episoden. Denn ChatGPT und künstliche Intelligenz werden gerade heiß debattiert. Wir werfen auf das Thema KI nochmal einen Blick aus einer neuen Perspektive. In der Folge vom 8. Februar. Zu sexy, wie Online-Algorithmen

Frauen benachteiligen. Eine junge Frau postet ein Bild bei Instagram. Sie will ihre Follower daran erinnern, regelmäßig zur Brustkrebsvorsorge zu gehen. Und Instagram löscht das Bild. Warum das Bild gelöscht wird, das klären wir in dieser Folge. Und wieso dahinter möglicherweise ein größeres Problem steckt. Nämlich Bilderkennungsalgorithmen. Und die können Frauen im Internet

systematisch benachteiligen. Ihr hört 11KM der Tagesschau-Podcast. Ein Thema in aller Tiefe, in der ARD-Audiothek. Heute mit der Datenjournalistin Katharina Brunner von BR-Data beim Bayerischen

Rundfunk. Ich bin Victoria Michalsack. Heute ist Mittwoch, der 8. Februar. Katharina, herzlich willkommen. Hallo, schön, dass ich da sein kann. Das Bild, das hat Rebecca Bauer gepostet. Rebecca Bauer ist eine Anfang 30-jährige Frau aus München. Und sie postet auf Instagram immer wieder zu, ich sag mal, Gesundheitsthemen, die Frauen betreffen. Da geht es dann um Instruktion, um Wechseljahre, um Indimetrie. Und eben auch um Brustkrebsvorsorge. Und im Oktober 2022 hat sie ein Bild von sich gepostet. Da stand sie daheim bei sich im Schlafzimmer, vorm Spiegel. Man sieht in der Ecke steht noch der Staubsauger. Und sie steht da oben ohne und hat ihre Hand so, ihren Arm so, über ihren Brüsten. Und sie hat ein Text dazu gepostet. Da stand drauf Monthly Reminder, Check Your Boobs, Hashtag Brustkrebsvorsorge.

Und hat das in eine Story gepostet und die wurde dann aber gelöscht. Ich hab mich ein bisschen angegriffen gefühlt. Also es war für mich einfach so, hey, grad ich, die sich da genau für einsetzt, dass ich eben nicht auf mein Körper reduziert werde oder ich nicht sexualisiert werde. Die bekommt dann so eine Meldung. Und da war ich auch etwas emperrt, muss ich sagen. Und der Grund, warum sie gesperrt wurde, von Instagram lautet sexuelle Kontaktaufnahme

durch Erwachsene. Sexuelle Kontaktaufnahme. Das war ja eigentlich nicht das, was damit gemeint war. Sie wollte ja eigentlich auf eine Brustkrebsvorsorge erinnern. Ganz genau. Also das heißt, Instagram geht davon aus, das ist irgendwie mehr oder weniger erotischer Inhalt oder irgendwie anzüglich erotisch gemeint. Genau. Wenn man sich dann durch die Richtlinien von Instagram durchkämmt, bis man zu diesem Punkt kommt, was das bedeutet, warum Bilder gesperrt werden, aus diesem Grund wegen der sexuell motivierten Kontaktaufnahme durch Erwachsene, dann gibt es zwei Kriterien, die ihr erfüllt sein müssen. Das erste ist Anbieten oder Nachfragen. Da kann man sagen, na gut, dieses Check Your Boobs ist da vermutlich irgendwie reingerutscht, dass dann Algorithmus automatisiert festgestellt hat, ja okay, das könnte irgendwie so ein Angebot sein. Ah, wegen Check, das ist dann vielleicht so ein bisschen wie wenn einer sagt, check mal Bio irgendwie für einen Link, wo ihr draufklickt oder check out this offer, also guckt euch dieses Angebot an, was ich euch mache, irgendwie so was. Genau. Und es braucht aber noch einen zweiten Punkt. Und da sind wir beim Thema, zwar braucht es anzügliche Elemente, zum Beispiel Posen. Und da scheint eben ihre Körperhaltung

oben ohne vom Spiegel, die eine Hand hält das Handy, die andere Hand ist sofort im Brüsten, das scheint da den Ausschlag gegeben zu haben. Ja, da merkt man schon, das kommt irgendwie nicht so ganz hin. Ich hätte das Bild jetzt zumindest nicht als anzüglich bewertet, also da war eigentlich gar nichts zu sehen. Die Pose fand ich jetzt auch nicht aufreizend, irgendwas muss da ja schief gelaufen sein. Ich frag mich ja, ob das bei einem Mann auch passiert wäre, mal so ganz generell. Das ist die Frage, die wir uns auch gestellt haben. Es geht um Algorithmen generell und die Frage, inwiefern sie in der Lage sind, solche schwammigen Kategorien wie sexuelle Anzüglichkeit denn korrekt zu bewerten? Also das heißt, du hast gerade schon gesagt, das ist natürlich eine automatische Erfassung, also da sitzt jetzt nicht einer irgendwo und muss jedes Foto auf Instagram oder auf Social Media generell durchkämmen, das wird ja gar nicht klappen, sondern das macht quasi eine künstliche Intelligenz. Genau, also wenn man was bei sozialen Plattformen postet zum Beispiel oder bei ganz vielen anderen

Anwendungen im Internet, laufen im Hintergrund Programme die Bilder prüfen. Zum Beispiel darauf, ob sie ebenso ansüßlich sind oder ob Gewalt darauf zu sehen ist oder pornografischer Materialien, Hakenkreuze, Alkohol, Drogen, was auch immer. Und Bilder werden geprüft und bekommen dadurch eine Bewertung, so einen Score, das ist das häufige Wert zwischen 0 und 1 beispielsweise. Und je nachdem, wie hoch dieser Wert ist oder ob mehrere Werte zusammenkommen, gibt es Regeln, die definiert sind, die wir von außen nicht kennen und ab irgendeinem Punkt springen die an und dann gibt es quasi eine automatische Entscheidung, ob ein Bild gezeigt werden darf oder eben nicht. Und das ist ja auch total sinnvoll, weil im Internet so viele Sachen gepostet werden den ganzen Tag. Dies ist eine richtige Flut, der man irgendwie her werden muss und ist ja gut, wenn keine Menschen sich solche, oft ja auch wirklich brutalen oder verstörenden Bilder ansehen muss. Also das ist ja grundsätzlich eine sinnvolle Angelegenheit. Die Frage ist halt, wo sind die Grenzen solcher Systeme? Und wir wollten wissen, wie funktionieren Algorithmen, die sagen, dass sie sexuelle Anzüglichkeit bewerten können.

Und wie habt ihr euch diese Algorithmen angeschaut in eurer Recherche? Was habt ihr da gemacht? Ich bin Teil eines Teams beim Bayerischen Rundfunk. Wir heißen BIA Data und wir sind Datenjournalistinnen.

Und das heißt, wir arbeiten in und mit Daten haben technische Fähigkeiten und wir schauen uns immer wieder solche technischen Systeme an, wie sie ja Teil unserer Lebenswelten sind. Wir sind ja überall davon umgeben. Sie sind auch immer irgendwie unsichtbar und deshalb schwer zu fassen.

Eigentlich machen wir was ganz Klassisch-journalistisches, nämlich wie beobachten. Nur halt in dem Fall nicht unbedingt Menschen oder politische Vorgänge oder ein Fußballspiel, sondern wir beobachten, was Algorithmen so tun. Man kann so ein Algorithmus, so ein Modell, der ja auch als Protagonisten versteht. Wir stellen eine Frage und kriegen eine Antwort.

Wie habt ihr das in diesem Fall gemacht? Wie ist diese Recherche abgelaufen?

Wir haben uns verschiedene Bilderkennungs-Software genommen, die es gibt, die es zu kaufen gibt, die kommerziell vertrieben werden und haben die getestet. Wir haben als Erstes recherchiert, welche Hersteller dann überhaupt solche Algorithmen anbieten. Da haben wir gesehen drei der fünf größten Unternehmen der Welt tun das. Google macht das, Microsoft, Amazon, aber es gibt

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

auch ein paar kleinere europäische Anbieter. Da haben wir uns Side-Engine angesehen. Ah, also auch schon eine große Bekannte dabei. Google, Microsoft und Amazon mit der Tochterfirma AWS.

Also die bieten alle so ein Standardreportoir an Bilderkennungssystemen an. Da gibt es dann ein Pornos und Gewalt oder halt sexuelle Anzüglichkeit. Google und Microsoft nennen das Ganze auf Englisch racy, ein bisschen seltsamer Begriff. Amazon und Side-Engine nennen das suggestive. Und wir sagen halt, das ist sexuell anzüglich. So fassen wir die vier Algorithmen zusammen. Das heißt, man kann sich da anmelden, Kritikartendaten hinterlegen und dann kann man quasi für einen kleinen Betrag pro Bild oder mit so einer Art Flatrate Bilder bewerten lassen.

Und das hat er dann gemacht. Ihr habt Bilder genommen und habt die Algorithmen bewerten lassen? Finden die das anzüglich ja oder nein?

Genau. Wir haben uns überlegt, welche Bilder wir brauchen, um testen zu können, ob es deutliche Unterschiede in der Bewertung von Frauen und Männern gibt. Und wir haben drei Themenbereiche abgedeckt.

Erstens, Bildern, auf denen Haut zu sehen sind, aber die aber vergleichsweise standardisiert fotografiert sind. Immer ähnliches Licht, die Leute machen nicht großartig, was da landet man schnell bei Unterwäschefotos, wie so Unterwäschekataloge oder Schwimmbekleidung. Standardisiert heißt einmal für Frauen, einmal für Männer? Oder wie habt ihr es gemacht?

Genau. Und wir haben uns quasi Fotos, zum Beispiel, wie man sie bei e-Commerce anbietet und findet die Unterwäsche verkaufen.

Also da geht es um die Unterwäsche, aber jetzt nicht so sexy lingerie Unterwäsche, sondern ganz einfach Katalogunterwäsche, wo die Leute jetzt auch nicht total anzüglich gucken oder so, sondern einfach Unterwäsche tragen.

Genau, sie haben nicht viel an und schauen in die Kamera. Und das Zweite sind Bilder, auf denen mit sexueller Anzüglichkeit gespielt wird. Solche Bilder wollten wir natürlich schon auch testen. Und Bilder, die in einem Umfeld aufgenommen worden sind, das per se nicht sexuell aufgeladen ist. Da geht es dann um Freizeit, um Sport, Beruf, solche Sachen. Und so haben wir mehr als 3000 Bilder gesammelt und die von jedem der Dienste bewerten lassen. Da schickt man quasi ein Bild, dann dauert es im Moment oder zwei und dann bekommt man Werte zurück. Bei Google einen Wert, bei Microsoft auch einen Wert, bei Amazon und Side-Engine gibt es mehrere Werte pro Bild. Und die haben wir alle gespeichert. Das heißt, wir haben 3000 Bilder und am Ende ungefähr 60.000 Bewertungen für diese ganzen Bilder und da muss man sich dann ein Reim drauf machen.

Dann habt ihr die ganzen Bilder gefüttert gehabt und die ganzen Zahlen bekommen. Wie war das, als das ausgewertet wurde?

Ich saß zu Hause im Homeoffice und habe Code geschrieben. Ich habe quasi diese Tabelle genommen mit den Bewertungen und ich habe durch ein statistisches Verfahren versucht herauszufinden, ob es denn deutliche Unterschiede in den Bewertungen gibt zwischen Männern und Frauen. Und das rechnet dann ein bisschen. Also das dauert dann so eine halbe, dreiviertel Stunde bis diese Berechnungen zu Ende sind und auch dann kommt eine Tabelle raus, die 3000 Zahlen lang ist. Die habe ich dann genommen und einen Balkendiagramm gemacht. Und zu gucken, in wie vielen Fällen finden wir so einen Gender-Bias, eine Verzerrung rein auf

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

Basis des Geschlechts und wann nicht. Und das war der Moment, als dann quasi bei allen vier Anbietern rote Balken erschienen sind. Also unsere Definition nach ein Gender-Bias besteht. Das war der Moment, wo klar ist, okay, da ist was dran. Das kann eine Geschichte sein.

Ah, okay. Also das heißt, ihr habt dann gesehen, da sind Unterschiede zwischen Frauen und Männern.

Die roten Balken, die zeigen, da ist ein fetter Unterschied.

Im Grunde haben wir zwei Sachen rausgefunden. Zum einen gibt es einen sogenannten Gender-Bias, eine Verzerrung nur aufgrund des Geschlechts. Das konnten wir bei allen vier Diensten nachweisen, unterschiedlich ausgeprägt und auch in unterschiedlichen Bereichen.

Also Frauen werden häufiger als anzüglich bewertet als Männer auf Fotos, auch wenn die eigentlich irgendwie das Gleiche machen.

Genau. Und das zweite Ergebnis ist, dass die Dienste für ein und dasselbe Bild immer wieder zu sehr abweichenden, sehr unterschiedlichen Ergebnissen kommen.

Ah.

Es gibt zum Beispiel ein Bild von einer Frau, die sitzt auf dem Sofa, einen großen roten Sofa, liest, hat eine Cheensan und ein gestreiftes T-Shirt, neben ihr sitzt so ein Hund und da sagt Google fünf von fünf, also sehr wahrscheinlich anzüglich. Und da fragt man sich schon als Betrachterin, warum.

Also das heißt, die kommen zu unterschiedlichen Ergebnissen, die Dienste untereinander oder derselbe Dienst, wenn man dem mehrfach das gleiche Bild zeigt.

Die Dienste untereinander.

Ah ja.

Das zeigt auch schon, wie schwierig es ist, so ein soziales Konzept wie sexuelle Anzüglichkeit irgendwie in so ein Modell zu gießen.

Das messbar zu machen quasi. Was erzeugt denn diese unterschiedliche Bewertung? Das habt ihr auch euch angeschaut, ne?

Genau. Wir wollten genauer verstehen, wie diese Algorithmen funktionieren und haben deshalb einen Test gemacht, der möglichst einfach ist. Und wir haben den zusammen mit der News WG vom Bayerischen Rundfunk gemacht.

Kennt ihr vielleicht, die machen Nachrichten bei Instagram und YouTube, Max Osenstetter ist einer der Hosts und mit ihm habt ihr einen Test gemacht.

Hallo.

Grüß dich.

Hallo.

Die hatten Lust mitzumachen bei unserem Test und der Max hat sich dafür oben ohne vorne Wand gestellt und hatte die Hände zu Beginn hinter dem Rücken verschränkt.

Und da hat er den BH, den sieht man nicht. Und den nimmt er dann so zur Seite, hält ihn in die Kamera und zieht ihn sich an. Und das ist dann so eine Sequenz von ungefähr zehn Sekunden und da haben wir so Serienbilder draus gemacht.

Also ganz viele Fotos.

Wir hatten dann ganz viele Fotos, 130 und die haben wir alle auch bewerten lassen, jedes einzelne. Und dann konnten wir richtig schön nachvollziehen, wie im Laufe dieser Bewegung,

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

die er macht, diese Scores für Anzüglichkeit hochgehen oder hochbleiben.

Okay, also das heißt, der hat sich den BH angezogen.

Und da kann man sehen, dass bei Microsoft und Side-Engine die Werte hochgehen, sobald der BH im Bild ist und vor allen Dingen, wenn er dann an seinem Körper ist.

Okay, also ich hätte jetzt glaube ich beide Bilder nicht als sonderlich anzüglich eingestuft, aber was hat denn die KI so gesagt?

Jetzt haben wir uns mal anschauen.

Ja, ich bin sehr gespannt.

Okay, jetzt bin ich wirklich aufgeregt.

Bei Microsoft war dieser Effekt super deutlich und auch bei dem Anbieter der Side-Engine war das so zu sehen.

Bei Amazon und Google war es anders, da war durchgehend eine hohe Bewertung.

Der Dienst von Microsoft sagt die Wahrscheinlichkeit dafür, dass dieses bildsexuelle Anzüglich ist, liegt bei 29 Prozent.

Okay, dann schauen wir mal, was passiert, wenn ich mir jetzt den BH überziehe.

Also, die Scala bleibt jetzt noch relativ gleich und jetzt ziehe ich ihn mir an und jetzt geht es wirklich mega krass nach oben, 99 Prozent mit BH.

Hä?

Also der BH macht den Unterschied.

Okay, also obwohl ich vorher weniger an hatte und man vorher meine Nippel gesehen hat und ich halt ja nackter war in dem Sinne und jetzt weniger nackt bin, sagt die KI.

Ja, also das ist zum 99 Prozent anzüglich.

Der BH macht den Unterschied und insbesondere bei Microsoft, bei denen wir auch den größten Gender Bias gefunden haben.

Okay, also Bilder von Frauen werden anders bewertet als die von Männern, aber was heißt denn das jetzt?

Das Problem ist, dass solche Bilderkennungssoftware in ganz vielen sozialen Plattformen und Apps und anderen Anwendungen im Hintergrund mitlaufen und immer wieder Bilder prüfen und selbst wenn es nur kleine Unterschiede immer gibt, dadurch, dass es so sehr viele Bewertungen gibt, jeden Tag an denen Dinge gepostet werden, kann es auch sein, dass solche kleineren Unterschiede

dann eine große Auswirkungen haben und solche Stereotypen am Ende dann ja auch immer weiter getragen werden.

Das nennt man dann Skalierungseffekt, den es bei Algorithmen generell immer wieder gibt.

Also das heißt, gibt es da konkrete Nachteile, die Frauen dann dadurch haben können?

Am Ende geht es um die Sichtbarkeit im Netz.

Was können wir sehen, was können wir nicht sehen, welche Bilder werden vielleicht gesperrt und welche werden vielleicht eher weniger gesperrt?

Und wenn gewisse Fotos nicht sichtbar sind, hat es natürlich Folgen für Aktivismus beispielsweise oder aber auch, wenn Leute Geld damit verdienen, wenn Leute diese Accounts nutzen zur Kundenarquise beispielsweise.

Ja, und das Foto von Rebecca, das war ja eigentlich dafür gedacht, dass Frauen daran erinnert

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

werden, zur Brustkrebsvorsorge zu gehen, das war dann ja auch weg.

Ganz genau.

Aber was läuft denn eigentlich falsch beim Algorithmus?

Woher kommt das denn?

Ja, solche Algorithmen, die fallen ja nicht vom Himmel, die sind ja Menschen gemacht.

Solche Modelle brauchen ganz viele Daten, mit denen sie trainiert werden können.

Das bedeutet, dass Menschen ganz viele Bilder bewertet haben im Vorfeld.

Da bekommt man Bilder vorgelegt und die müssen Fragen beantworten, zum Beispiel ist da eine sexualanzügliche Pose darauf zu sehen oder Unterwäsche, ein Dekotee, ein Bikini, ein Mann mit freiem Oberkörper, welche Fragen auch immer, so genau können wir das von außen nicht wissen.

Und all diese Informationen werden dann zusammengekommen und sind die Basis für ein Modell. Also Menschen sagen dem Algorithmus erst mal, was sie als anzüglich bewerten und dann kann der Algorithmus bei anderen Bildern, die man ihm zeigt, ausspucken, wie anzüglich ist das wohl.

Genau.

Also entweder was die Menschen bewerten, was sie selber als anzüglich halten oder man stellt quasi so indirekte Fragen, aus denen man dann schließen kann, dass das wahrscheinlich anzüglich sein könnte, zum Beispiel Bikinis oder Dekoties.

Ist natürlich total schwierig, weil das heißt, der Algorithmus spiegelt ja nur wieder, was Menschen bewerten an Anzüglichkeit und Anzüglichkeit ist ja so ein wahnsinnig schwammiger Begriff.

Genau, Anzüglichkeit ist Ansichtssache.

Wenn man sich mit dem Thema beschäftigt, dann kommt man früher oder später auf den obersten Gerichtshof in den USA, da gab es in den 60er Jahren eine Debatte darüber, was da ein Obstdröckchen sei.

Und der Richter meinte damals, I know it when I see it.

Und es ist so eine Formulierung, die benutzt man jetzt immer, wenn etwas ja total subjektiv ist und keine klar definierten Parameter ist.

Und in dieser Kategorie spielt halt auch die sexuelle Anzüglichkeit.

Also ich weiß das, wenn ich sehe, aber ich kann ja nicht erklären, auf welcher Grundlage ich das dann entscheide, wenn ich sehe.

Man kann es nicht klar definieren, immer und oft nur so, bisschen von hinten durch die Brust ins Auge.

Und das muss man dann irgendwie den Algorithmen beibringen.

Ja, eben.

Also wenn nicht mal Menschen klar sagen können, wie sie das machen, dann ist es natürlich umso schwieriger, das eine künstliche Intelligenz machen zu lassen oder fast unmöglich vielleicht sogar.

Oder dass es klar ist, dass es einfach Unschärfen gibt und Bereiche, in denen das Ganze dann halt einfach nicht mehr so gut funktioniert.

Das können wir aber von außen wiederum nicht wirklich nachvollziehen, wo die Grenzen der einzelnen Systeme liegen, der einzelnen Dienste, weil die Hersteller eben kaum detaillierte

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

Informationen dazu rausgeben.

Die Hersteller, das sind ja Google, Microsoft und Amazon, mit dem Tochterunternehmen AWS, also Amazon Web Services, was sagen die denn dazu?

Ja, die haben wir alle kontaktiert und gefragt und die sagen alle, naja, wir berechnen lediglich Wahrscheinlichkeiten.

Und was dann mit den Bewertungen der einzelnen Bilder passiert, das müssen dann die Kunden entscheiden, also zum Beispiel soziale Plattformen, irgendwelche Apps oder Webseiten, das müssen die dann selbst entscheiden.

Seit Engine und Amazon sagen auch, sie bieten verschiedene Unterkategorien an, um dann verschiedene

Anwendungsfälle abzubilden.

Und Google meint auch, man könne keine Hinweise auf systemische Probleme finden, aber Zitat kein Filter ist zu 100 Prozent genau.

Wichtig ist zu wissen, dass unsere Recherche keine generelle Aussage über diese Algorithmen treffen kann.

Unsere Ergebnisse gelten immer nur für unsere Test, also für unsere 3000 Bilder mit unseren verschiedenen Kategorien, wo wir uns sorgfältig überlegt haben, welche Bilder wir da reinnehmen oder nicht.

Aber es bietet Hinweise darauf.

Das ist das Problem ganz grundsätzlich, wenn man versucht, solche Manentes im Fachjargau audits zu machen, also Algorithmen von außen zu bewerten und zu verstehen, wie dann die überhaupt funktionieren.

Und das Problem ist, wenn man in eine Apotheke geht und man sich Kopfschmerz tabletten kauft, dann gibt es ein By-Pack-Zettel.

Da steht eng bedruckt drauf, was die Risiken und Nebenwirkungen und auch Wechselwirkungen der Medikamente sind.

So etwas in der Art bräuchte es auch für solche Modelle, damit eben Leute, die das einsetzen, die sich diesen Dienst kaufen oder anderweitig Interesse daran haben, eben auch wissen, was sind die Risiken, was sind die Nebenwirkungen, was sind die Grenzfälle, wo funktionieren diese Systeme einfach nicht mehr richtig gut.

Und wo funktionieren sie schon gut, weil diese Fälle gibt es ja natürlich schon auch.

Ja, wenn du das so sagst, auch was die Gefahren sind, denke ich so ein bisschen an ganz andere Algorithmen und künstliche Intelligenzen, bei denen es ja auch wichtig ist zu merken, wie funktionieren die und wie geben die das wieder, was wir den füttern.

Ja genau, und Verzerrungen, die findet man ja überall.

Wir haben mit einer Expertin gesprochen, sie heißt Abeber, Birhane und die kennt sich mit solchen Sachen sehr gut aus und die meinte man findet die Algorithmen immer und auch unserem anderen Experten, der war auch überhaupt nicht überrascht, wenn es um Gesichtserkennung

geht beispielsweise oder auch solche Sachen wie ChatGPT oder generative Bildsysteme oder KIS, die Bilder generieren, man findet überall Verzerrungen, nicht nur was Geschlecht betrifft, auch Ethnie, Alter und so weiter.

Das ist Abeber, Birhane, sie ist Kognitionswissenschaftlerin an der Uni von Dublin.

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

Sie sagt also, dass es kein Datensatz ohne Verzerrung gibt, weil wenn man Verzerrungen komplett loswerden will, dann müsste es eine Sicht auf die Welt geben, die frei ist von irgendwelchen menschlichen Einschätzungen.

Das ist frei von human, human view, aber das ist impossible, weil everybody will disagree on how I should label this data set.

Wir genau sehen diese Verzerrungen denn aus.

Also diese Verzerrungen können sich dann je nach dem, um was es sich handelt, auch unterschiedlich äußern.

Es gab Untersuchungen zu Gesichtserkennungen, wo schwarze Frauen beispielsweise nicht gut erkannt wurden, was daran liegt, dass zu wenige Trainingsdaten von schwarzen Frauen vorhanden waren.

Und das System, das halt einfach nicht gut war und das hat man erst rausgefunden, als Wissenschaftlerinnen sich das genauer angesehen haben, so ähnlich wie wir und Tests gemacht haben.

Im schlimmsten Fall automatisieren solche Systeme also Stereotype.

Sie erstellen Modelle, die nicht präzise sind, die weit von der Realität entfernt und am Ende schädlich sind.

Das sagt Birhane da.

Also das heißt, KIs und Algorithmen sind am Ende immer nur so schlau wie schlau, wie sie füttern.

Und darauf müssen wir achten.

Genau.

Wie schlau man sie füttert und wie gut man sie kalibriert.

Wenn wir jetzt schauen, was könnten denn da Lösungen sein?

Im Grunde gibt es zwei Lösungen.

Das eine ist die Systeme zu verbessern, indem man an die Trainingsdaten rangeht.

Es geht darum, dass da alle verschiedene Gruppen gleich behandelt werden und die Modelle gut kalibriert werden.

Und das andere ist, ich glaube, es braucht mehr Transparenz, weil die Hersteller müssen ja nichts deklarieren und so bildet sich irgendwie so eine Art Kaskade der Intransparenz.

Die Hersteller sagen nicht genau, wie ihre Systeme funktionieren.

Die Kunden müssen dann nehmen, was kommt und das setzen sie dann einfach wiederum ein, ohne dass die Nutzerinnen und Nutzer genaueres wissen.

Für Nutzerinnen wie der Beckerbauer ist es dann quasi mehrfach intransparent.

Richtlinien sind ja natürlich dafür da, dass man sich nach ihnen richtet und dass man darauf aufpasst, dass man nicht dagegen verstößt, aber dann möchte ich auch nicht nur so Pseudorichtlinien vorgesetzt bekommen, die halt auch vielleicht doppeldeutig sind oder eben nicht ganz klar, sondern ich möchte ganz klar sagen, das ist erotischer Inhalt oder eben auch zu sagen, ja, Frauen und Männer, sie werden einfach unterschiedlich beurteilt und ich weiß, sie dürfen es nicht sagen, weil es wieder gegen dieses Diskriminierungsgesetz geht, aber sie machen es einfach.

Jetzt haben wir noch mal von der Rebecca gesprochen, die dieses Bild auf Instagram gepostet hat,

[Transcript] 11KM: der tagesschau-Podcast / Zu sexy - wie Online-Algorithmen Frauen benachteiligen (Wiederholung)

dass man zur Brustkrebsvorsorge gehen soll und das gelöscht wurde, aber bei eurer Untersuchung war Instagram ja nicht dabei, also ihr habt jetzt aber mit eurer Untersuchung nicht rausfinden können, warum genau das so ist oder wie so das bei Instagram so funktioniert.

Nein, das konnten wir leider nicht rausfinden, es hätten wir gerne gemacht, Meta der Konzern, zu dem ihr Instagram gehört, der stellt seine Technologien dann nicht zahlenden Kunden zur Verfügung, sondern die benutzen das nur selbst.

Wir haben aber natürlich auch bei Meta nachgefragt, was es denn mit Rebecca Bowers Bild und der Sperrung auf sich hat und nachdem wir uns bei ihnen gemeldet haben, ist eine interessante Sache passiert, nämlich das Bild wurde wieder freigeschaltet.

Ach, aber kommentarlos.

Genau.

Ja gut.

Paustest du jetzt anders Bilder auf Insta und Co oder wie betrachtest du das jetzt?

Na ja, es ist klar, dass diese Systeme, die sind einfach längst in unserem Alltag und in unserer Lebenswelten angekommen, wir sind ja überall umgeben davon, deshalb geht es nicht darum, ob wir sie einsetzen, sondern wie wir sie einsetzen, unter welchen Umständen.

Katarina, danke, dass du uns davon erzählt hast.

Sehr gerne.

Katarina Brunner hat das Thema recherchiert, gemeinsam mit ihren Kolleginnen Elisah Harlan und Shannon Reitmeier von BR Data.

Mehr Infos dazu findet ihr auf br24.de.

Da findet ihr auch ein interaktives Tool von BR Data, das sie programmiert haben.

Da könnt ihr selbst testen, ob eure Definition von Anzüglichkeit mit der von den Algorithmen übereinstimmt.

Also ihr bewertet ein Bild mit einem Schieberegler und dann bekommt ihr die Einschätzung der Algorithmen.

Uns könnt ihr gerne ein Abo in der ARD-Audiothek dalassen, dann verpasst ihr keine Folge.

Folgenautorin ist Mira-Sophie Potten, mitgearbeitet hat Marc Hoffmann, Produktion Viktor Werresch,

Gerhard Wichow, Eva Erhardt, Jürgen Kopp und Simon Schuling, Redaktionsleitung, Fumiko Lipp und Lena Gürtler.

Mein Name ist Victoria Michalsack.

FKM, der Tagesschau-Podcast, wird produziert von BR24 und NDR Info.

Wir hören uns morgen wieder, ciao.