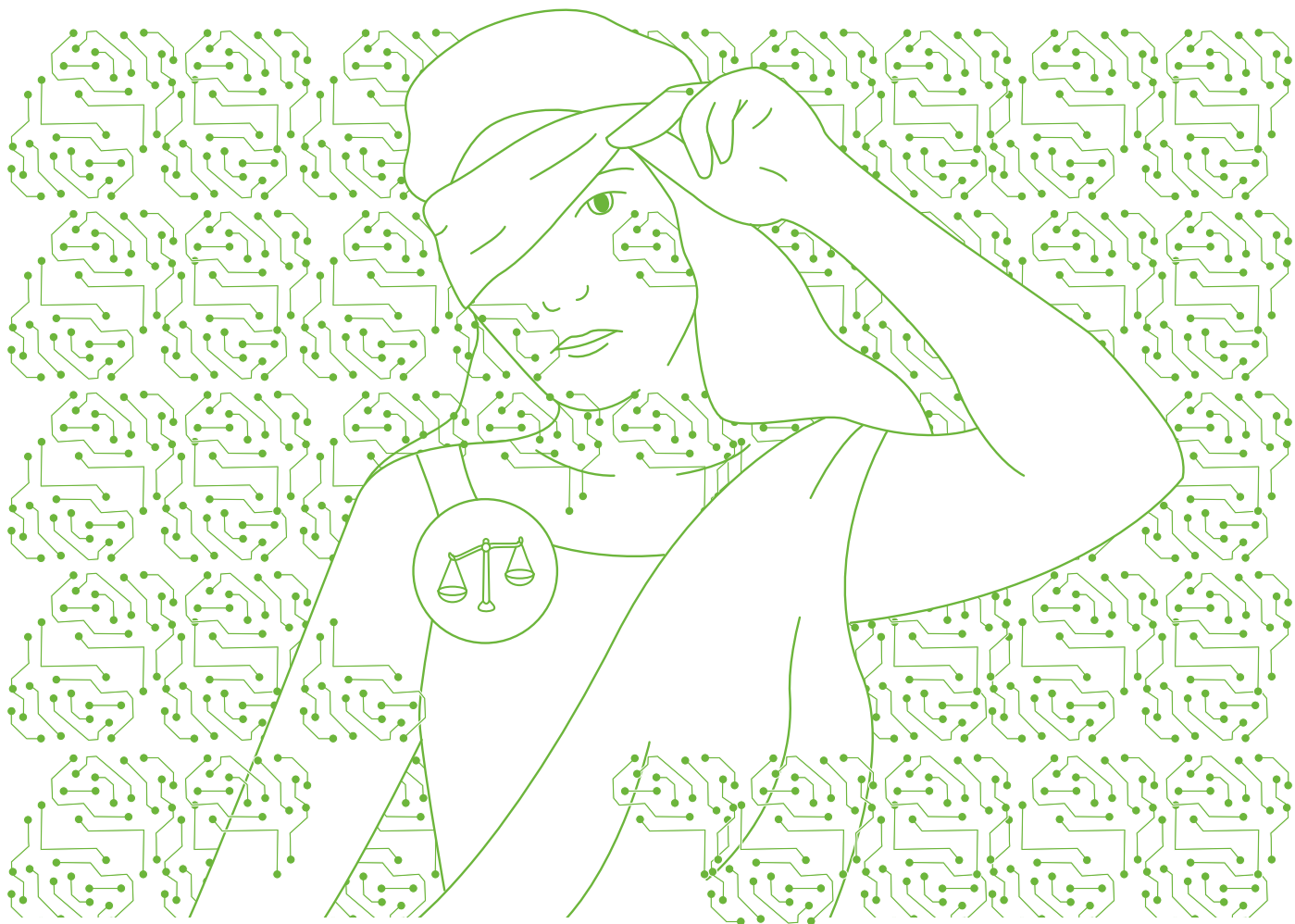


Künstliche Intelligenz ist nicht diskriminierender als die Daten, mit denen sie gefüttert wird

Interview mit **Henrike Schlottmann**
Jan Kuhlen



Der Einsatz künstlicher Intelligenz (KI) als nächste Stufe algorithmenbasierter Entscheidungen breitet sich branchenübergreifend aus und damit auch das Risiko für Diskriminierung von Einzelpersonen und Gruppen. Henrike Schlottmann und Jan Kuhlen vom «Zentrum für digitalen Fortschritt» erklären im Interview, was unternommen werden kann, um unerwünschte Ungleichbehandlung zu reduzieren. Im Zentrum ihrer Einschätzung steht die These, dass künstliche Intelligenz gesellschaftliche Verzerrungen lediglich widerspiegelt und dass durch die Systematisierung der Fehler die Diskriminierung einfacher korrigierbar werden kann – im digitalen wie im analogen Leben.

Warum ist künstliche Intelligenz diskriminierend?

Jan Kuhlen: Diskriminierung meint die Gleichbehandlung von Ungleichen oder die Ungleichbehandlung von Gleichem - und zwar ohne Rechtfertigung. Grundsätzlich ist Technologie nicht diskriminierend, ausser der Mensch zieht diskriminierende Schlüsse aus ihr, oder schult sie so, dass die Ergebnisse eine Diskriminierung automatisch nach sich ziehen. Dasselbe gilt für künstliche Intelligenz. Ungewollte Diskriminierung kann bei ihrem Einsatz entstehen, weil unklar ist, wie ein künstliches neuronales Netz auf seine Ergebnisse kommt. Es fehlt aktuell ein Bewusstsein für die massgeblichen Kriterien zur Beurteilung der Herausforderungen.

Welche Formen der Diskriminierung sind beim Einsatz von künstlicher Intelligenz bekannt?

Henrike Schlottmann: Die Diskriminierung kann auf vier verschiedenen Ebenen stattfinden: Sie hängt von den Trainingsdaten, den gewählten Algorithmen, dem Team und den Zielvorgaben ab. Es ist der Mensch, der sich für diskriminierende Zielvorgaben entscheidet oder der Teams einsetzt, die bereits blinde Flecken haben. Die Diskriminierung basiert in der Regel auch beim Einsatz von KI auf den geschützten Attributen wie Geschlecht, Hautfarbe oder Religion. Im Prinzip handelt es sich um die gleiche Problematik wie im analogen Leben, denn der mit Daten gefütterte Algorithmus spiegelt die Verzerrung - auch Bias genannt - lediglich wider. Diese besteht indes bereits in der Gesellschaft.

Können Sie ein Beispiel geben?

Henrike Schlottmann: Ein klassisches Beispiel ist der Recruiting-Algorithmus bei Amazon¹, der beim Bewerbungsverfahren diskriminierend wirkte: Er kam zum Schluss, dass Bewerbungen von Frauen schlechter zu bewerten seien. Dies, weil der Algorithmus auf einem unausgeglichene Datenset trainiert wurde - und zwar anhand von Bewerbungen, die innerhalb der letzten zehn Jahre bei Amazon eingegangen waren. Die meisten davon kamen von Männern. Ein Muster, aus dem das System folgerte, dass Männer die bevorzugten Kandidaten waren. Ein weiteres Beispiel ist das «Predictive Policing» - also vorausschauende Polizeiarbeit, bei der automatisierte Entscheidungen in die Strafverfolgung einbezogen werden. So wurden in den USA Vorhersagen zu Rückfallquoten mit einem inhärenten Bias in den Trainingsdaten gemacht – und Nicht-Weisse auf Basis ihrer Hautfarbe systematisch diskriminiert ².

1: Künstliche Intelligenz diskriminiert (noch): <https://www.zeit.de/arbeit/2018-10/bewerbungsroboter-kuenstliche-intelligenz-amazon-frauen-diskriminierung>

2: Predictive Policing: Dem Verbrechen der Zukunft auf der Spur: <https://www.bpb.de/dialog/netzdebatte/238995/predictive-policing-dem-verbrechen-der-zukunft-auf-der-spur?type=galerie&show=image&i=238997>

Verstehen Sie unter «rationaler Diskriminierung» denn auch die Optimierung des Individuums auf Kosten des Solidargefüges, wie es der Soziologe Stefan Selke definiert?

Jan Kuhlen: Das «Life-Logging», also das Sammeln grosser Datenmengen aus dem täglichen Leben eines Individuums, führt zu einer stärkeren Vermessung und Bewertung des Menschen. Dies könnte auch eine stärkere Vermessung der Gesellschaft zur Folge haben. Der Mensch kann dadurch in einem grösseren Ausmass mit anderen verglichen werden. Diese Vermessung, so die Kritik, raubt uns ein Stück Privatheit.

Henrike Schlottmann: «Rational» trägt immer auch etwas Systematisches in sich. In Bezug auf die Diskriminierung bedeutet dies, dass der Bias in der Gesellschaft bereits besteht und vom Algorithmus widergespiegelt wird, durch die Anwendung wird das Ganze aber systematisiert. Es gibt dann nicht mehr den Menschen, der situationsbedingt entscheidet, sondern das System, das systematisch agiert. Gleichzeitig wird die Diskriminierung dadurch sichtbarer und korrigierbarer.

Gibt es Anwendungen, bei denen besonders stark rational diskriminiert wird?

Jan Kuhlen: Bei Gesundheitsdaten, die durch Apps gesammelt werden, ist die Gefahr der rationalen Diskriminierung besonders gross. Für die Diskriminierung durch künstliche Intelligenz kann man das jedoch nicht grundsätzlich sagen. Da ist die Gefahr von Diskriminierung vor allem dann vorhanden, wenn es um die Bewertung von Personen geht - insbesondere bei den geschützten Attributen.

In welchem Verhältnis stehen «Life-Logging» und der Einsatz künstlicher Intelligenz?

Henrike Schlottmann: «Life-Logging» oder Big Data alleine ist noch keine künstliche Intelligenz, bildet aber die Voraussetzung für intelligente Analysen.

Jan Kuhlen: Künstliche Intelligenz, wenn man sie denn so einsetzt, kann aufgrund dieser Daten Ableitungen und Generalisierungen machen, denen keine Kausalität zu Grunde liegt. Zum Beispiel: Wer jeden Morgen um 7 Uhr aufsteht und drei Kilometer joggt, ist erfolgreicher.

Gibt es rationale Diskriminierung auch im analogen Leben?

Jan Kuhlen: Ja. Wenn die Versicherung mir einen Fragebogen zuschickt, dann fällt die Risikoauswertung auch anhand dieser Daten aus. Die Auswertungen werden durch Big Data aber feingliedriger und können mithilfe von künstlicher Intelligenz leichter systematisiert werden, was auch zu besseren Ergebnissen führt: Sie ermöglichen eine adäquate Berücksichtigung der individuellen Risikosituation. Das ganze System wird durch künstliche Intelligenz gerade massiv aufgedeckt. So hat man auch die rationale Diskriminierung bisher zu wenig wahrgenommen.

Sie fordern in einem Positionspapier³ verbindliche Regeln für den Einsatz von KI, um Missbrauch zu verhindern.

Henrike Schlottmann: Wir fordern einen globalen Ethikrat, der sicherstellt, dass bestimmte Regeln und Guidelines für alle gelten.

Wie könnte dadurch Diskriminierung verhindert werden?

Jan Kuhlen: Grundsätzlich muss versucht werden, auf den vier genannten Ebenen Diskriminierung zu verhindern: Das Team muss so ausgesucht werden, dass keine blinden Flecken bestehen, die Daten müssen in Lern- und Überprüfungsdaten aufgeteilt und im Zweifel anhand von Black-Box-Analysen kontrolliert werden - einem indirekten Verfahren, das bei-

3: Der Einfluss Künstlicher Intelligenz auf Freiheit, Gerechtigkeit und Solidarität: <https://d-64.org/wp-content/uploads/2018/11/D64-Grundwerte-KI.pdf>

spielsweise durch Visualisierung nachträglich Erklärungsmodelle herbeiführt. Auch wenn die Mechanik einer algorithmischen Black-Box - also eines geschlossenen Systems - unbekannt ist, kann man das Verhalten, also Ein- und Ausgaben, prüfen.

Henrike Schlottmann: Die Systeme müssen möglichst nachvollziehbar und erklärbar sein. Man sollte verstehen können, wie die Entscheidungen zustande kommen. Die Algorithmen könnten beispielsweise einem Audit unterzogen werden; die Daten würden dann auf Bias untersucht. Die dafür nötigen Instrumente hängen aber immer von der persönlichen Definition von Fairness ab und sind noch lange nicht fertig optimiert.

Müssen wir überhaupt immer mit Kausalitäten argumentieren können, um Entscheidungen plausibel zu begründen?

Jan Kuhlen: Bei staatlichen Entscheidungen, denen wir uns nicht entziehen können, erwarten wir kausale Argumentationen, die transparent und nachvollziehbar sind. Solche Entscheidungen sollten nicht anhand einer Wahrscheinlichkeit erfolgen. Künstliche Intelligenz aber basiert immer auf Wahrscheinlichkeitswerten. Entsprechend ist beim Einsatz von KI Vorsicht geboten. Wir können von einer Technologie jedoch nicht grundsätzlich erwarten, dass eine Eingabe immer eine komplett nachvollziehbare Ausgabe ergibt. Von Lebewesen erwarten wir das auch nicht – nicht einmal von Menschen. Die Gesellschaft wird sich daran gewöhnen müssen, dass Technologie in Laborzuständen nicht hundertprozentig kausale Ergebnisse liefert und trotzdem in der Gesamtschau deutlich präziser sein kann als menschliches Handeln.

Welche Einsatzgebiete von KI sind denn legitim?

Jan Kuhlen: Bereits heute gibt es zahlreiche Technologien mit automatisierten Lernkomponenten, die gut funktionieren: Übersetzungen, Videoanalysen oder die Spracherkennung. Die Nachvollziehbarkeit, also wie ein Algorithmus zu guten Übersetzungsergebnissen kommt, dürfte die Wenigsten interessieren. Wichtig ist, dass der Satz am Ende gut übersetzt ist. Auch bei der Spracherkennung - zum Beispiel Spracheingaben, die in Text verwandelt werden - ist es für die meisten Nutzer irrelevant, warum etwas erkannt wird.

Aber warum kann man nicht argumentieren, dass Technologien, die nicht nachvollziehbar sind, komplett verboten werden müssten, weil sie - ohne Rechtfertigung - immer diskriminierend wirken könnten?

Jan Kuhlen: Würde man die Frage auf Menschen anwenden, müssten wir auch die Entscheidungen von Menschen einschränken, weil sie ohne Rechtfertigung immer diskriminierend wirken könnten. Künstliche Intelligenz ist grundsätzlich eine Errungenschaft mit enormen Potenzialen. Aber wie gesagt: In manchen Gebieten, zum Beispiel dort, wo staatliche Entscheidungen getroffen werden, muss man ihren Einsatz genauer prüfen.

Die Einführung eines globalen Ethikrates scheint illusorisch, da normative Grundsätze nicht weltweit gelten, wie auch das Online-Spiel «Moral Machine» des Massachusetts Institute of Technology (MIT) deutlich gemacht hatte⁴: Probanden wurden mit Unfallsituationen selbstfahrender Autos und variierenden Verkehrsteilnehmern konfrontiert und mussten sich für den in ihren Augen vertretbaren Unfallschaden entscheiden. Das Resultat: ethische Prinzipien variieren geografisch stark.

Jan Kuhlen: Gewisse Regeln oder Normen haben sich weltweit durchgesetzt. Auch im Bereich der künstlichen Intelligenz halte ich das nicht für ausgeschlossen. Zum Beispiel wird die Ächtung autonomer Waffensysteme bereits im Koalitionsvertrag zwischen CDU, CSU

4: Moral Machine: Kann ein selbstfahrendes Auto ethisch handeln?: <https://www.dw.com/de/moral-machine-kann-ein-selbstfahrendes-auto-ethisch-handeln/a-46045294-0>

und SPD gefordert ⁵. Dass das – wie bei Atomwaffen – nur weltweit funktionieren kann, liegt auf der Hand. Genauso dürfte man global zur Einsicht kommen, dass bei Bots oder Avataren kenntlich gemacht werden müsste, dass es sich dabei um keine Menschen, sondern um Maschinen handelt. Auch muss geklärt werden, wann Entscheidungen erklärbar und nachvollziehbar sein müssen. Natürlich wird es umso schwieriger, einen gemeinsamen ethischen Wertekodex zu finden, je mehr Länder und Kulturen man versucht einzubeziehen.

Was passiert, wenn wir es nicht schaffen?

Henrike Schlottmann: Wie die Welt aussehen könnte, wenn die USA oder China jeweils ihre eigenen Regeln aufstellen, lässt sich erahnen, wenn wir das sogenannte Social Credit System betrachten. Im Reich der Mitte wird derzeit versucht, eine Überwachungstechnik aufzubauen, die jeden Einzelnen auch in einer unüberschaubaren Masse sofort identifizieren kann. Wer sich gut verhält, wird belohnt, für schlechtes Verhalten gibt es Strafen. Da stellt sich die Frage: Ist das langfristig die Welt, in der wir leben wollen? Ich finde, es sollte der Anspruch sein, bestimmte Regeln international zu diskutieren, weil künstliche Intelligenz nun einmal ein globales Thema ist. Dessen Anwendungen stehen rund um den Erdball zur Verfügung, und das Internet ermöglicht allen Menschen Zugriff auf diesen Markt.

Sie fordern zudem einen solidarischen Umgang mit Daten. Was verstehen Sie darunter?

Jan Kuhlen: Es geht darum, sinnvollen Mehrwert aus den Daten zu ziehen. Das heisst, dass nicht-personenbezogene Daten der Gesellschaft - solidarisch - zur Verfügung gestellt werden sollten, um daraus soziale und am Gemeinwohl orientierte Innovationen entstehen zu lassen. So können Vereinfachungen durch künstliche Intelligenz auch dem Service public nutzen. Heute scheitern viele Ideen noch an dem nicht vorhandenen Datenmaterial oder der fehlenden Aufbereitung, die viel Zeit und Ressourcen verschlingt. So hat die deutsche Wirtschaft aktuell Probleme, genügend qualitativ hochwertiges Datenmaterial zu erhalten. Gleichzeitig ist viel Wissen über deutsches Konsumverhalten bei den amerikanischen Tech-Riesen wie Facebook oder Google hinterlegt. In der deutschen Wirtschaft ist zwar auch Datenmaterial vorhanden, das wird aber nicht geteilt und deckt jeweils nur einen spezifischeren Teil ab.

Was schlagen Sie vor?

Jan Kuhlen: Man könnte zum Beispiel intelligente Datenplattformen schaffen, welche die Daten trainieren oder tauschen. Dafür werden gemeinsame Standards benötigt, damit die verschiedenen Akteure ihre Daten auf eine Plattform hochladen und anderen - gegen Geld oder kostenlos - zur Verfügung stellen können.

Henrike Schlottmann: Dabei muss natürlich sichergestellt werden, dass die Daten nicht personalisiert sind, ansonsten könnten bestimmte Daten bestimmten Personen zugeordnet werden. Und nicht jeder Mensch möchte, dass Informationen über die eigene Person frei verfügbar sind. Daten zu entpersonalisieren ist schwierig, weil es nicht reicht, nur den Namen zu entfernen. Es muss verhindert werden, dass aus den anderen Attributen wie Strasse oder Ausbildung trianguliert werden kann, um welche Person es sich handelt.

Fordern Sie auch einen solidarischen Umgang mit Gesundheitsdaten?

Jan Kuhlen: Gesundheitsdaten zu sammeln und zu teilen sehe ich als problematisch, weil sie ein hochprivates Wissen in sich tragen. Es geht um nicht weniger als um den Gesundheitszustand eines Menschen. Nicht umsonst gibt es eine ärztliche Schweigepflicht. Es ist gesellschaftlich anerkannt, dass dieses Wissen nicht geteilt werden muss. Ganz grundsätz-

5: Koalitionsvertrag zwischen CDU, CSU und SPD: <https://www.bundesregierung.de/breg-de/themen/koalitionsvertrag-zwischen-cdu-csu-und-spd-195906>

lich geht es aber ohnehin nicht ums Teilen von personenbezogenen Daten, sondern eben von entpersonalisierten. Gleichzeitig ist uns wichtig, dass die Datensysteme nicht angegriffen werden, sonst hätten wir ein Problem mit der persönlichen Selbstbestimmung. Jeder braucht denselben Gesundheitsschutz, egal ob er morgens um 7 Uhr joggen geht oder nicht. Die gesetzlichen Krankenkassen dürfen keine algorithmisch begründeten Unterschiede machen, sei es in Bezug auf das Leistungsspektrum oder die Kosten.

Solidarität ist ein menschliches Konzept. Was bedeutet das für die Entwicklung von KI-Systemen?

Jan Kuhlen: Technik ist grundsätzlich neutral, kann aber natürlich auch solidarisch genutzt werden. Wenn durch KI beispielsweise die Stromnetze besser verwaltet werden und dadurch der Stromverbrauch sinkt, dann käme das allen Menschen zugute. Alle Nutzungen, die sich am Gemeinwohl orientieren, können und sollen gefördert werden.

Henrike Schlottmann: Am Ende sind es auch politische Entscheidungen. Dabei ist klar: Wenn ich eine unsolidarische Zielvorgabe habe, wird das Resultat kein solidarisches sein. Wenn ein Algorithmus beispielsweise genutzt wird, um die Sportlichkeit oder das Körpergewicht zu ermitteln und in der Konsequenz ein schlechterer Krankenversicherungstarif angeboten wird, dann ist das unsolidarisch. KI kann also nur solidarisch sein, wenn Menschen mit dem Algorithmus auch solidarische Zielvorgaben verfolgen.

Abschliessend: Welchen gesellschaftlichen Nutzen sehen Sie durch KI?

Jan Kuhlen: Der negativen Darstellung von künstlicher Intelligenz ist entschieden entgegen zu treten. Der grosse Erfolg ist, dass Prozesse automatisiert und einfache Arbeiten reduziert werden können. Denn: Dies schafft viel Platz für neue Aufgaben und Kreativität. Wir müssen uns bewusst machen, dass die digitale Revolution eine industrielle Revolution ist.

Henrike Schlottmann: Ich möchte gerne nochmals auf die Diskriminierung zurückkommen, denn auch hier hat KI ihre positiven Seiten: Die Algorithmen zeigen eine gewisse Diskriminierung auf, wie sie in der Gesellschaft gelebt wird, und dank der Systematisierung kann man aktiv dagegen vorgehen.

Henrike Schlottmann und **Jan Kuhlen** leiten die Arbeitsgemeinschaft Künstliche Intelligenz bei D64 – Zentrum für digitalen Fortschritt, einem gemeinnützigen digitalpolitischen Verein. Schlottmann hält einen MSci in Mathematik vom University College London. Sie war mehrere Jahre bei einer internationalen Managementberatung tätig und hat dort Projekte im Bereich Digitalisierung und Innovation begleitet. Heute unterstützt sie als Co-CEO von ProjectTogether Social Entrepreneurs bei der Umsetzung ihrer Projekte. Kuhlen ist studierter Jurist und Soziologe. Er berät mit seiner Kanzlei KUHLEN Partnerschaft von Rechtsanwälten mbB Startups, Agenturen und mittelständische Unternehmen in wirtschaftsrechtlichen Fragen. Jan Kuhlen ist Mitglied der Enquete Kommission Künstliche Intelligenz des deutschen Bundestages und des Expertenrates Künstliche Intelligenz.
