Diskriminierungsrisiko: Vom Algorithmus benachteiligt



Studie des KIT zeigt vielfältige Möglichkeiten der Ungleichbehandlung ? und empfiehlt präventive Maßnahmen

Nicht nur Unternehmen, auch staatliche Institutionen setzen verstärkt auf automatisierte Entscheidungen durch algorithmenbasierte Systeme. Ihre Effizienz spart Zeit und Geld - birgt jedoch vielfältige Gefahren der Benachteiligung einzelner Menschen und ganzer Bevölkerungsgruppen. Zu diesem Ergebnis kommt eine jetzt vorgestellte Studie der Technikfolgenabschätzung des Karlsruher Instituts für Technologie (KIT) im Auftrag der Antidiskriminierungsstelle des Bundes. Ob bei der Kreditvergabe, der Auswahl neuer Mitarbeitender oder bei juristischen Entscheidungen - in immer mehr Bereichen werden Algorithmen dazu eingesetzt, menschliche Entscheidungen vorzubereiten oder sie ihnen gleich ganz abzunehmen. "Dass dies zwangsläufig zu objektiveren und damit faireren Entscheidungen führt, erweist sich heute leider oft als Trugschluss", so Carsten Orwat, vom Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des KIT. "Kritisch wird es insbesondere dann, wenn die Algorithmen mit tendenziösen Daten arbeiten und auf eigentlich geschützte Merkmale zurückgreifen", so der Autor. Hierzu gehören vor allem Alter, Geschlecht, ethnische Herkunft, Religion, sexuelle Orientierung oder Behinderungen.

Welche Ursachen diese Formen der Diskriminierung haben, wie sie sich auf die Gesellschaft auswirken und sich Diskriminierungsrisiken künftig verringern lassen, hat Carsten Orwat im Auftrag der Antidiskriminierungsstelle des Bundes eingehend untersucht. Die Studie "Diskriminierungsrisiken durch Verwendung von Algorithmen" veranschaulicht anhand von 47 Beispielen, wie Algorithmen auf unterschiedliche Weise diskriminieren können und wie man dies nachweisen kann.

Immobilien, Kredite, Justiz und mehr: Vielfältige Beispiele für Diskriminierungsrisiken

Der Autor der Studie beschreibt beispielsweise Vorgänge auf dem Immobilien- und Kreditmarkt oder im Strafvollzug. In den USA gebe es mehrere dokumentierte Fälle, in denen die Algorithmen Sozialer Medien Wohnungsanzeigen zugelassen hätten, die durch den 'Fair Housing Act' geschützte Personengruppen nicht zu sehen bekamen - etwa Migranten, Menschen mit Behinderung oder Nicht-Weißer Hautfarbe, so der Studienautor. In Finnland wurde ein Kreditinstitut zu Strafzahlungen verurteilt, weil sein Algorithmus bei der automatisierten Online-Kreditvergabe Männer gegenüber Frauen und finnische gegenüber schwedischen Muttersprachlern bei der Kreditvergabe benachteiligte. Eine Ungleichbehandlung, die nach dem finnischen Antidiskriminierungsrecht verboten ist. US-amerikanische Richter arbeiten bei der Entscheidung über vorzeitige Haftentlassungen mit einem umstrittenen System, das Risikoscores berechnet. Journalisten- und Menschrechtsverbände kritisieren, dass das System das Rückfallrisiko von Schwarzen Menschen systematisch zu hoch bewertet.

"Bei Systemen des maschinellen Lernens wird es häufig problematisch, wenn KI-Systeme mit Daten trainiert werden, die Ungleichbehandlungen oder Stereotypen abbilden", erklärt Carsten Orwat. "Dann spiegeln auch die so erzeugten Algorithmen die Ungleichbehandlungen oder Stereotypen wider. Werden Daten verarbeitet, die Bewertungen von Menschen über anderer Menschen

beinhalten, so können sich Ungleichheiten und Diskriminierungen sogar verbreiten oder verstärken." Das zeigte sich beispielsweise in den USA bei einem System für Lebensmittel- und Gesundheitskontrollen, das auf diskriminierenden Bewertungen von Restaurants basierte.

Empfehlungen zur Regulierung

Die Gesellschaft muss diese Ungleichbehandlungen aber nicht tatenlos hinnehmen. Die Studie nennt mehrere Ansatzmöglichkeiten, um Diskriminierungen bei algorithmenbasierten Differenzierungen zu begegnen. "Am sinnvollsten erscheinen präventive Maßnahmen", sagt Carsten Orwat. So könnten Firmen ihre Personal bzw. IT-Mitarbeiter von Antidiskriminierungsstellen beraten lassen. Diese Angebote könnten auch dazu sensibilisieren, nur solche Datensätze zu verwenden, die keine diskriminierenden Praktiken oder Ungleichheiten widerspiegeln.

Das Ziel sei es, so Orwat, dass Algorithmen in Zukunft "diskriminierungsfrei by design" werden. Dazu müssten Programme bereits während ihrer Entwicklung geprüft werden. Letztlich gehe es dabei immer um den Schutz von gesellschaftlichen Werten wie Gleichheitsziele oder Schutz der freien Entfaltung der Persönlichkeit. Damit diese auch angesichts der rasanten Entwicklungen von "Big Data" und KI gewährt bleiben, sei es an verschiedenen Stellen nötig, das Antidiskriminierungs- und Datenschutzrecht zu verbessern.

Die vollständige Studie als PDF zum Download:

https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Expertisen/Studie_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.html