

Diskriminierung – trotz oder wegen künstlicher Intelligenz?

Vielfältige Entscheidungen, die konkret die Lebensführung und Entwicklungschancen von Menschen betreffen, werden seit Langem regel- und merkmalsbasiert getroffen. Mediziner und Kostenträger berücksichtigen Erkrankungs- und Rückfallrisiken von Patienten, wenn sie über Behandlungen oder Rehabilitationen entscheiden. Versicherungen klassifizieren ihre Kunden mit Blick auf Schadenswahrscheinlichkeiten und berechnen entsprechende Tarife, Banken treffen Kreditentscheidungen auf Basis von Einschätzungen der Kreditwürdigkeit beispielsweise durch die SCHUFA Holding AG.

Seitdem diese Entscheidungen zunehmend datenbasiert und mittels Algorithmen/Software automatisiert werden (algorithmische Entscheidungssysteme [AES]), rücken diese in den Fokus der öffentlichen Wahrnehmung, und von Wissenschaft, Politik und anderen Stakeholdern werden damit verbundene Chancen und Risiken thematisiert. AES bieten das Potenzial, Entscheidungen schneller, effizienter und objektiver zu treffen und Unzulänglichkeiten menschlicher Entscheidungen vermindern oder gar vermeiden zu können.

Gleichwohl gehen direkte Risiken automatisierter Entscheidungen mit Fehlentscheidungen und den daraus erwachsenden Folgen sowie oftmals unklaren Verantwortungs- und Haftungsfragen einher. Bisher gibt es keine spezifischen Offenlegungspflichten, weder für eingesetzte (Trainings-)Daten, anhand derer Kriterien für Entscheidungsregeln und -modelle abgeleitet werden, noch für die Algorithmen, mit denen diese Regeln auf neue Fälle angewendet werden. Dadurch sind Entscheidungsprozesse von außen oft nicht nachvollziehbar. Die fehlende Transparenz bereitet den Nährboden für Misstrauen oder Ablehnung. Durch die zunehmende Anwendung maschineller Lernverfahren wird diese Situation verstärkt und auf eine neue Stufe gehoben.

Seit einiger Zeit wird darauf hingewiesen, dass jenseits von Fehlentscheidungs-

risiken AES dazu führen können, dass einzelne Personen oder Personengruppen systematisch diskriminiert werden und dass dies aufgrund der fehlenden Transparenz nur schwer zu erkennen und nachzuweisen ist. Der Vorwurf der möglichen Diskriminierung wiegt gesellschaftlich schwer, weil Grundrechte von Einzelnen betroffen sind. In Deutschland besteht ein Diskriminierungsverbot nicht nur für staatliche Stellen, sondern auch für den privaten Bereich. Einzelne Fälle von AES, die Diskriminierung Vorschub leisteten, sind bereits öffentlich geworden (wobei nicht immer klar ist, inwiefern tatsächlich Menschen geschädigt wurden).

Ziel und Vorgehensweise

Die Diskriminierungsrisiken durch AES stellen nicht nur eine technische, sondern eine juristische, gesellschaftliche und politische Herausforderung dar. Zentrale Fragen sind insbesondere:

- In welchen Bereichen und für welche Entscheidungen werden bereits algorithmische Entscheidungssysteme eingesetzt, die nach geltender Rechtslage diskriminierungsfrei sein sollten (insb. gemäß Allgemeinem Gleichbehandlungsgesetz [AGG])? In welchen Bereichen ist eine Anwendung von AES und maschinellem Lernen in Zukunft wahrscheinlich? Welche Vorteile erwartet man sich von deren Einsatz?
- Welche Risiken von Diskriminierung entstehen durch den Einsatz von AES? Welche Fälle von Diskriminierung aufgrund von algorithmischen Entscheidungssystemen sind bereits bekannt geworden?
- Nach welchen Kriterien und Verfahren sollte der behördliche Einsatz algorithmischer Entscheidungssysteme zugelassen werden? Welche behördlichen Anwendungen gibt es bereits, welche sind möglich?
- Welche technischen Möglichkeiten existieren, um Diskriminierung durch algorithmische Entscheidungssysteme

und maschinelles Lernen im Vorhinein oder nachträglich zu erkennen und zu beheben?

- Welche Verfahren (wie z. B. counterfactual explanations) werden in der Literatur diskutiert, um die Nachvollziehbarkeit komplexer Verfahren herzustellen? Wie können dabei Geschäftsgeheimnisse gewahrt bleiben?
- Welche rechtlichen Möglichkeiten und Vorschläge zur Einhegung von Diskriminierungsrisiken durch Algorithmen gibt es, beispielsweise im Rahmen des Wettbewerbsrechts oder durch behördliche Kontrollmechanismen? Liegen diesbezüglich bereits Erfahrungen in anderen Ländern vor?

Das TAB wird die Vielzahl vorliegender und erwarteter Studienergebnisse anhand dieser Leitfragen auswerten und in Form einer Synopse zusammenfassen, um einen sachlich fundierten Überblick über die noch junge Debatte zu schaffen. Ergänzend sollen Interviews mit Expertinnen und Experten geführt werden, um mögliche Widersprüche bzw. Unschärfen zu klären.

TA-Projekt

Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen

Themeninitiative

Ausschuss für Bildung, Forschung und Technikfolgenabschätzung

Kontakt

Dr. Arnold Sauter
+49 30 28491-110
sauter@tab-beim-bundestag.de