Transparente Algorithmen – Wie lässt sich algorithmische Diskriminierung verhindern?

Entscheidungen werden zunehmend durch Algorithmen vorbereitet oder automatisch getroffen. Dies gilt für selbstfahrende Autos, vernetzte Haushaltsgeräte, Entscheidungsprozesse im Gesundheitswesen oder die Mediennutzung. Auch Einkaufsempfehlungen durch digitale Sprach-Assistenten wie Alexa, Siri, Cortana und Co. sowie die Bewertung von Menschen in der Arbeitswelt, bei Versicherungen oder Banken beruhen auf algorithmischen Entscheidungssystemen und selbstlernenden Programmen (siehe Thema Zukunft der Bewertungsplattformen). In vielen Fällen liefern Algorithmen die Grundlage für Entscheidungen, die von existenzieller Bedeutung sind. Wer einen Kredit erhält und zu welchen Konditionen dieser vergeben wird (siehe Thema Fintechs), ob die Aufnahme in eine Versicherung möglich ist und wie hoch die zu zahlende Prämie ist, gehört ebenso dazu wie Algorithmen, die vorschlagen, wer zu einem Vorstellungsgespräch eingeladen werden, befördert oder entlassen werden soll (vgl. Schaar 2017). Da das zentrale Merkmal der auf Big Data basierenden algorithmischen Steuerung die Klassifizierung ist, d. h. die Zuordnung von Datenelementen zu bestimmten Gruppen, ergibt sich ein hohes Diskriminierungspotential.[1] Die Diskriminierung von BewerberInnen aufgrund von Geschlecht, ethnischer Zugehörigkeit oder Religion ist verboten, die Auswahl nach Zugehörigkeit in sozialen Netzwerken wie Facebook oder LinkedIn dagegen nicht (Boyd et al. 2014). So greifen Algorithmen in den Alltag und die Autonomie der Einzelnen ein, ohne dass diese Möglichkeiten haben, die Entscheidungen nachzuvollziehen oder eben diesen Entscheidungen zu widersprechen. Algorithmische Systeme wenden ihre Entscheidungslogik konsistent auf alle Fälle an, unterliegen keiner subjektiven Verzerrung, diskriminieren damit auch konsistent, wie sich in Bezug auf Bewerbungen bereits zeigt (Boyd et al. 2014). Hinzu kommt, dass sich Algorithmen auf komplexe Situationen einstellen, also selbst im Wandel sind.[2]

[1] Aktuell zeigt sich dies an der Diskussion um den so genannten AMS-Algorithmus in Österreich. Das Arbeitsmarkt-Chancen-Assistenzsystem (AMAS) sollte ab 2021 auf Basis von Statistiken vergangener Jahre die zukünftigen Chancen von Arbeitssuchenden am Arbeitsmarkt berechnen. Die Arbeitssuchenden werden dabei anhand der Prognose ihrer „Integrationschance“ in drei Gruppen eingeteilt, denen unterschiedliche Ressourcen für Weiterbildung zugeteilt werden. Wie eine Studie des ITA (Allhutter et al. 2020) zeigt, hat der AMS-Algorithmus weitreichende Konsequenzen für Arbeitssuchende, AMS-MitarbeiterInnen sowie die Organisation AMS.

[2] So kann eine künstliche Intelligenz stereotype Werturteile entwickeln, wenn sie ihr Wissen aus repräsentativen Texten der Menschheit generiert, die eben diese Stereotype enthalten, siehe Caliskan, et al. (2017). Für eine Vielzahl von Beispielen, wie Meinungen und unhinterfragte Hypothesen in mathematische Modelle eingebettet sind, siehe O'Neil (2016).

Für einige Bereiche wie das Gesundheitswesen oder die Versicherungsbranche stellen sich konkrete Fragen nach der Notwendigkeit neuer gesetzlicher Regelungen und Kontrollmöglichkeiten. Wenn Geschäftsmodelle in Zukunft stark auf algorithmischen Systemen beruhen, deren Funktion den Kern des Geschäftsmodells ausmachen, können Transparenz und Kontrolle im Widerspruch zur unternehmerischen Freiheit sowie der Eigentumsfreiheit im Hinblick auf die algorithmischen Systeme stehen. Die seit Mai 2018 geltende EU-Datenschutz-Grundverordnung kann einzelfallbezogene Transparenz, Nachvollziehbarkeit und Korrigierbarkeit von automatisierten Entscheidung ermöglichen, ist jedoch auf individuelle Rechte bezogen und eignet sich daher nicht für die Analyse und Regulierung von gruppenbezogenen, systematischen Diskriminierungsrisiken und soziotechnischer Risiken (vgl. Dreyer/Schulz 2017; Goodman/Flaxman 2017). Gegenwärtig entsteht ein Forschungsbereich unter dem Namen „Diskriminierungsbewusstes Data Mining“, der erforscht, wie Vorhersagemodelle frei von Diskriminierung erstellt werden können, v. a. wenn die historischen Daten, auf denen sie aufgebaut sind, möglicherweise verzerrt oder unvollständig sind oder sogar diskriminierende Entscheidungen aus der Vergangenheit enthalten (Žliobaitė 2017).

Dieser Technologiebereich hat eine hohe Innovationsdynamik, sodass ein analytischer und politischer Zugang zu algorithmischen Entscheidungsprozessen notwendig ist, der die zukünftigen Potentiale und disruptiven gesellschaftlichen Veränderungen von algorithmischen Entscheidungssystemen als Ausgangspunkt nimmt.[3] Auf grundlegender Ebene stellt sich die Frage, welche politischen Herausforderungen die intransparenten Systeme und das abzusehende Maschinenlernen aufwerfen und wie Transparenz, ethische Bewertungskriterien und demokratische Kontrolle gewährleistet werden können.[4] Algorithmen können auch Kompromisse zwischen konkurrierenden Werten transparent machen, was bedeutet, dass Algorithmen nicht nur eine Bedrohung darstellen, die reguliert werden muss, sondern mit den richtigen Schutzmaßnahmen einen potenziell positiven Beitrag zu Gerechtigkeit leisten können (Kleinberg et al. 2019). Da sie auch in zentralen Infrastrukturen der Zukunft steuernd wirken werden, geht es darum, Möglichkeiten der Evaluierung, Transparenz und Kontrolle auch über die gesetzliche Regulierung hinaus zu untersuchen. Die diskriminierenden Folgen der Klassifikation durch Algorithmen stehen dem Diskriminierungsverbot entgegen und können verfassungsgesetzlich gewährleistete Rechte in hohem Maße tangieren. Eine vorausschauende Befassung muss insbesondere die internationale Dimension der Technologieentwicklung berücksichtigen, die nur begrenzt gesetzlich zu regeln ist.

[3] Mittlerweile hat die EU-Kommission einen Vorschlag zur Regulierung von KI vorgelegt, siehe digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[4] Siehe dazu auch Udrea et al. 2022.

Zitierte Quellen

Allhutter, D., Mager, A., Cech, F., Fischer, F. und Grill, G., 2020, *Der AMS-Algorithmus – Eine soziotechnische Analyse des Arbeitsmarkt-chancen-Assistenz-Systems (AMAS)*, November 2020, Wien: ITA.

Boyd, D., Levy, K. und Marwick, A., 2014, The networked nature of algorithmic discrimination, *Data and discrimination: Collected essays*.

Caliskan, A., Bryson, J. J. und Narayanan, A., 2017, Semantics derived automatically from language corpora contain human-like biases, *Science 356(6334)*, 183-186.

Dreyer, S. und Schulz, W., 2017, *Was bringt die Datenschutz-Grundverordnung für automatisierte Entscheidungssysteme? Potenziale und Grenzen der Absicherung individueller, gruppenbezogener und gesellschaftlicher Interessen,* Gütersloh: Bertelsmann Stiftung.

Goodman, B. und Flaxman, S., 2017, European Union Regulations on Algorithmic Decision Making and a “Right to Explanation”, *Ai Magazine 38(3)*, 50-57.

O'Neil, C., 2016, *Weapons of math destruction: How big data increases inequality and threatens democracy:* Crown.

Udrea, T., Fuchs, D., & Peissl, W. (2022). Künstliche Intelligenz. Verstehbarkeit und Transparenz – Endbericht (p. 77). Wien. doi:/10.1553/ITA-pb-2022-01

Schaar, P., 2017, Überwachung, Algorithmen und Selbstbestimmung, in: bpb: Bundeszentrale für politische Bildung (Hg.): *Digitale Gesellschaft und politisches Handeln*, Bonn.

Kleinberg, J., Ludwig, J., Mullainathan, S. und Sunstein, C. R., 2019, Discrimination in the Age of Algorithms, *Journal of Legal Analysis 10*, 113-174.

Žliobaitė, I., 2017, Measuring discrimination in algorithmic decision making, *Data Mining and Knowledge Discovery 31(4)*, 1060-1089.