AUTOMATISIERUNG

Künstliche Intelligenz kann diskriminieren

Viele Verbraucher betrachten es bis heute mit Skepsis, wenn Entscheidungen allein von Algorithmen getroffen werden – zu Recht, wie ein im Mai veröffentlichtes Forschungspapier des Leibniz-Instituts für Finanzmarktforschung SAFE zeigt, das auf Basis von Daten im Zuge eines Experiments, das in den Jahren 2016 bis 2019 mit mehr als 3600 Personen durchgeführt wurde: Die Gefahr, bestimmte Gruppen von Verbrauchern durch den Einsatz von KI systematisch zu benachteiligen, besteht tatsächlich.

Das liegt an der Natur solcher Systeme: Um nämlich die Algorithmen des KI-Systems gezielt einsetzen zu können, müssen sie mit großen Datensätzen trainiert werden. Diese Daten können jedoch systematische Verzerrungen für Personen einer bestimmten Gruppe enthalten, zum Beispiel aufgrund ihres Geschlechts, Einkommens, Bildungsstandes oder Alters. „Die historisch systematische Benachteiligung bestimmter Bevölkerungsgruppen kann sich im Zeitalter intelligenter Algorithmen potenzieren und zunehmend unerwünschte gesellschaftliche und ökonomische Konsequenzen haben", so Kevin Bauer, einer der Autoren der Studie.

Wurden beispielsweise bei der Vergabe von Krediten durch Banken Daten mit einem unterproportionalen Anteil an Frauen gesammelt, kann das die Prognosesicherheit einer eingesetzten KI beeinträchtigen: Der auf diese Daten trainierte Algorithmus wäre dann systematisch schlechter darin, die Kreditwürdigkeit für Frauen adäquat zu bestimmen. Der Algorithmus kann somit Gefahr laufen, im Durchschnitt eine geringere Wahrscheinlichkeit für Frauen bei der Rückzahlung des Kredits vorherzusagen. Ein so automatisiertes KI-System würde Frauen dann seltener Kreditwürdigkeit zuschreiben.

Damit würde das mit unzureichenden Daten von Frauen trainierte KI-System eine gesellschaftliche Ungleichbehandlung verstärken.

Eben das ist der Grund, weshalb die Schufa – entgegen immer wieder vorgebrachter Vorwürfe – kein Geoscoring auf Basis der Wohnanschrift durchführt: Kunden sollen keinen schlechteren Scorewert erhalten, nur weil andere Verbraucher aus der gleichen Wohngegend sich als weniger kreditwürdig erwiesen haben oder einfach weniger häufig in der Datenbasis vorkommen.

Ein anderes Beispiel ist die biometrische Identifikation anhand der Gesichtserkennung. Hier hat sich gezeigt: Wenn die Datenbasis zu wenige Menschen mit dunkler Hautfarbe umfasst, kann das dazu führen, dass Dunkelhäutige häufiger nicht zuverlässig erkannt werden.

Erfahrungen mit der Benachteiligung bestimmter Gruppen durch vermeintlich neutrale Algorithmen haben auch Unternehmen gemacht, die im Personalbereich die Vorauswahl aus einer großen Bewerberzahl automatisieren: Wenn beispielsweise bestimmte Positionen in der Vergangenheit überwiegend mit Männern oder kaum mit Menschen mit Migrationshintergrund besetzt worden waren, suchte das System auch unter neuen Bewerbern bevorzugt Männer und Menschen ohne Migrationshintergrund aus, da diese anscheinend für die entsprechende Position besonders geeignet waren.

Das Bewusstsein dafür vorausgesetzt, lassen sich solche Probleme natürlich deutlich vermindern und gegebenenfalls sogar lösen, indem kontinuierlich Feedbackschleifen mit repräsentativen, nicht verzerrten Daten eingebaut werden. Dann lernt der Algorithmus fortlaufend weiter, bis er die sogenannte algorithmische Diskriminierung überwindet und dann tatsächlich „vorurteilsfrei" funktioniert.

Das wiederum heißt: „Um den Erfolg des KI-Systems zu messen, ist es nach wie vor nötig, dass Menschen die Leistung und damit die Qualität eines trainierten Algorithmus überwachen", so Bauer. Red.