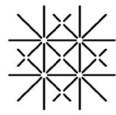




Universität
Zürich ^{UZH}



Universität
Basel

Juristische
Fakultät

WHITE PAPER

Diskriminierung

Florent Thouvenin
Stephanie Volz

Juni 2024

CENTER FOR
INFORMATION
TECHNOLOGY
SOCIETY AND
LAW — ITSL

e-PIAF

electronic Public Institutions and
Administrations Research Forum

Dieses White Paper wurde im Projekt «**Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz**» entwickelt, das vom Center for Information Technology, Society, and Law (ITSL) der Universität Zürich und von der Forschungsstelle electronic Public Institutions and Administrations Research Forum (e-PIAF) der Universität Basel durchgeführt und von der Stiftung Mercator finanziell unterstützt wird. Dieses White Paper ist Teil einer Reihe von White Papers, die sich mit den zentralen Herausforderungen befassen, die mit dem Einsatz von Künstlicher Intelligenz (KI) in Unternehmen und in der Verwaltung verbunden sind.

Die White Papers und weitere Informationen zum Projekt finden sich auf www.itsl.uzh.ch und www.ius.unibas.ch/e-piaf.

Folgende White Papers sind bislang erschienen:

- **Manipulation**
- **Diskriminierung**
- **Datenschutz**
- **Transparenz durch Begründung von Verfügungen**
- **Transparenz durch öffentliches Verzeichnis**

Das Kernprojektteam besteht aus folgenden Personen:

Prof. Dr. Florent Thouvenin, Professor für Informations- und Kommunikationsrecht an der Universität Zürich, Vorsitzender des Lenkungsausschusses des ITSL

Prof. Dr. Nadja Braun Binder, MBA, Professorin für öffentliches Recht an der Universität Basel

Dr. Stephanie Volz, wissenschaftliche Geschäftsführerin des ITSL und Lehrbeauftragte an der Universität Zürich

Liliane Obrecht, MLaw, Wissenschaftliche Mitarbeiterin und Doktorandin an der Juristischen Fakultät der Universität Basel

Systeme Künstlicher Intelligenz (KI) können – wie Menschen – diskriminieren. Ursachen für Diskriminierung finden sich bei der Entwicklung, in den Daten und in der Anwendung. Die Bundesverfassung verbietet Diskriminierung durch den Staat, Diskriminierung durch Private wird in der Schweiz aber bisher nur fragmentarisch rechtlich erfasst. Das breite Anwendungsfeld von KI-Systemen lässt Diskriminierung durch Private zunehmend problematisch und eine rechtliche Erfassung notwendig erscheinen. Mögliche Lösungen sind die Einführung eines allgemeinen Gleichbehandlungsgesetzes, die Statuierung eines neuen Grundsatzes der Datenbearbeitung im Datenschutzgesetz (DSG) und die Erfassung als Persönlichkeitsverletzung über das allgemeine Persönlichkeitsrecht (Art. 28 ZGB).

Problemstellung

KI-Systeme treffen zunehmend Entscheidungen mit relevanten Wirkungen für unser Privat- und Berufsleben. Diese Entscheidungen sind problematisch, wenn Menschen aufgrund von geschützten Merkmalen wie z. B. Herkunft, Alter oder Geschlecht unterschiedlich behandelt werden, ohne dass dafür ein sachlicher Grund besteht. **Diskriminierende Entscheidungen** wurden bisher meist von Menschen getroffen. Diskriminierung ist daher keine KI-spezifische, sondern eine **gesamtgemeinschaftliche Problematik**. Beim Einsatz von KI-Systemen bestehen aber besondere Herausforderungen.

Der Einsatz von KI-Systemen führt in aller Regel zu einer **Skalierung**: Die Systeme sind darauf ausgerichtet, massenhaft Entscheidungen zu treffen und umzusetzen; diskriminierende Entscheidungen von KI-Systemen wirken sich deshalb regelmässig auf eine Vielzahl von Personen aus. Hinzu kommt, dass bestehende Diskriminierungen beim Einsatz von KI-Systemen aufgrund von **Rückkoppelungseffekten** («feedback

loops») weiter verstärkt werden können. Zudem lässt sich Diskriminierung durch KI-Systeme oft nur schwer erkennen und nachweisen. Grund dafür ist die oft **mangelnde Nachvollziehbarkeit** der Empfehlungen und Entscheidungen solcher Systeme, insb. bei selbstlernenden Systemen, die je nach Komplexität eine eigentliche «Black Box» sein können.

Der Einsatz von KI-Systemen kann zwar zu Diskriminierungen führen, diese Systeme können aber auch dazu beitragen, **Diskriminierungen zu verhindern**. Denn anders als bei menschlichen Entscheidungen ist es bei Entscheidungen von KI-Systemen grundsätzlich möglich, bestehende Diskriminierungen nicht nur zu erkennen, sondern durch eine Anpassung der Systeme auch zu beheben. Das mag technisch anspruchsvoll sein und nicht immer vollständig gelingen. Im Vergleich zum gesellschaftlichen Prozess, der erforderlich ist, um Diskriminierungen durch Menschen zu verhindern, ist die Anpassung von KI-Systemen aber ungleich einfacher, schneller und kostengünstiger.

Phänomen und Begriff

Der Begriff der Diskriminierung kann unterschiedlich verstanden werden. In diesem White Paper wird der rechtlich definierte Begriff der Diskriminierung im Sinn des **Diskriminierungsverbots der Bundesverfassung** (Art. 8 Abs 2 BV) verwendet. Dieser bezeichnet die Ungleichbehandlung von Personen aufgrund eines geschützten Merkmals ohne qualifizierte Rechtfertigung. Eine qualifizierte Rechtfertigung liegt vor, wenn die Ungleichbehandlung gestützt auf einen sachlichen Grund erfolgt und verhältnismässig ist. Im Recht hat der Begriff der Diskriminierung stets einen negativen Gehalt.

Im Zusammenhang mit Diskriminierung durch KI-Systeme werden regelmässig die Begriffe «(Algorithmic) Bias» und «(Algorithmic) Fairness» verwendet. «Bias» kann eine wertende Bedeutung wie Voreingenommenheit oder Vorurteil haben, aber auch wertneutral als Verzerrung oder Abweichung von einem Standard verstanden werden. «(Algorithmic) Fairness» wird oft mit Diskriminierungsfreiheit gleichgesetzt. Die Begriffe sind jedoch nicht deckungsgleich, weil sich «Fairness» nicht auf Ungleichbehandlungen gestützt auf geschützte Merkmale beschränkt.

Zu unterscheiden ist zwischen **direkter Diskriminierung**, die eine Ungleichbehandlung direkt auf ein geschütztes Merkmal stützt und **indirekter Diskriminierung**, die vorliegt, wenn eine an sich neutrale Regelung, d.h. eine Regelung, die nicht an ein geschütztes Merkmal anknüpft, in ihren tatsächlichen Auswirkungen Angehörige einer geschützten Gruppe besonders benachteiligt. Indirekte Diskriminierung ist nur rechtlich relevant, wenn sie von erheblicher Bedeutung ist.

Im Zusammenhang mit KI-Systemen gibt es verschiedene Erscheinungsformen von Diskriminierung, die aufgrund der Charakteristika solcher Systeme besonders relevant sind:

– Im Vordergrund steht die Problematik der «Proxies» (auch: Stellvertretervariablen). Bei der **Proxy-Diskriminierung** hängt eine Entscheidung von Daten ab, die keine geschützten Merkmale enthalten, die aber mehr oder weniger stark mit einem geschützten Merkmal korrelieren. Proxy-Diskriminierungen treten auf, weil KI-Systeme nach Korrelationen suchen, und es viele Merkmale gibt, die mit geschützten Merkmalen korrelieren. So kann bspw. die Zugehörigkeit zu einer Gruppe in Social Media ein Proxy für das Geschlecht

oder die sexuelle Orientierung sein. Proxy-Diskriminierung entsteht in der Regel ungewollt und ist für Aussenstehende und oft auch für die Entwickler:innen selbst nur schwer zu erkennen.

– Auch die **intersektionelle Diskriminierung** kann bei KI-Systemen eine relevante Rolle spielen. Entscheide von KI-Systemen beruhen in der Regel auf einer Kombination von mehreren Merkmalen. Wenn für eine Entscheidung mehrere geschützte Merkmale relevant waren, ist oft schwierig nachzuvollziehen, welcher Faktor den grössten Einfluss hatte.

Ursachen von Diskriminierung bei KI-Systemen

Die Ursachen für Diskriminierung können in verschiedenen Phasen des Lebenszyklus von KI-Systemen liegen, von der Entwicklung und dem Training bis zur Anwendung. In diesem White Paper können nur einige der wichtigsten Ursachen genannt werden. Dabei ist zu beachten, dass sich die verschiedenen Kategorien und Fallgruppen nicht immer klar trennen lassen und voneinander abhängen können.

Entwicklung

Der Ursprung einer Diskriminierung durch KI-Systeme liegt häufig in der Entwicklung und Programmierung. Die Auswahl der Architektur und die Abläufe bzw. der Code können absichtlich oder unabsichtlich Verzerrungen enthalten. Bei der Entwicklung von KI-Systemen sind verschiedenen Entscheidungen zu treffen, die sich diskriminierend auswirken können, so etwa:

– **Architektur und Design:** Bereits in der Architektur und im Design eines KI-Systems, bspw. in der Festlegung des Trainingsverfahrens für ein Modell, können Ursachen für Diskriminierung liegen. Ein KI-System wird selten von Grund auf neu entwickelt; vielmehr werden meist zahlreiche bestehende Komponenten anderer Systeme integriert, bspw. Teile von Programmcodes oder Datenbanken. Diese stammen oft aus einem anderen Kontext und enthalten Wertungen von verschiedenen Personen, die Diskriminierungen enthalten können.

– **Datenauswahl:** Die Person, welche die Trainingsdaten auswählt, entscheidet, welche Daten für das Training des Modells verwendet werden. Es besteht die Gefahr, dass Menschen Trainingsdaten von Personen wählen, die ihnen ähnlich sind, bspw. bezogen auf

Geschlecht, Alter und Hautfarbe. Dieser Effekt kann verstärkt werden, wenn KI-Systeme von Personen mit ähnlichem gesellschaftlichem Hintergrund programmiert werden.

Trainieren, Validieren und Testen

Datensätze, die für das Trainieren, Validieren und Testen eines KI-Systems verwendet werden, können aus verschiedenen Gründen **unausgewogen** sein. Beim Training sind vor allem die folgenden Konstellationen relevant:

- **Historical Bias:** Daten sind immer ein Abbild der Vergangenheit. Sie können deshalb auf veralteten, diskriminierenden Realitäten beruhen. Gesellschaftliche Veränderungen werden möglicherweise nicht abgebildet. Dadurch können Probleme der Vergangenheit weiter reproduziert werden.
- **Representation Bias:** Häufig repräsentiert ein Datensatz nicht alle Personengruppen ausgewogen. Einerseits gibt es für gewisse Personengruppen derzeit weniger Daten als für andere. Dies kann historisch bedingt sein, z.B. weil früher medizinische Versuche überwiegend an weissen Männern gemacht wurden. Umgekehrt gibt es für gewisse Gruppen überproportional viele Daten, etwa wenn bei der vorausschauenden Polizeiarbeit (predictive policing) Gebiete, in denen bereits viel Kriminalität beobachtet wurde, aufgrund von Datenanalysen noch stärker kontrolliert und deshalb noch mehr Kriminalfälle registriert werden. Auch wie und wo Daten gesammelt werden, kann über die Ausgewogenheit der Trainingsdaten entscheiden: Werden Daten bspw. im Internet erhoben, sind weibliche Personen in der Regel unterrepräsentiert. Ein Bias kann auch entstehen, wenn KI-Systeme mit Daten über Personen trainiert werden, die nicht repräsentativ für die Personen sind, auf welche die Systeme angewendet werden. Bspw. können Daten von Personen mit Wohnsitz in den USA unter Umständen ungeeignet sein, um KI-Systeme zu trainieren, die auf Personen in der Schweiz angewendet werden sollen.

Bei der Aufbereitung der Trainingsdaten für ein selbstlernendes KI-System werden die Trainingsdaten teilweise mit einem **Label** versehen, um die Daten zu kennzeichnen. Wenn ein KI-System bspw. Katzen erkennen soll, wird es mit Bildern trainiert, die mit «Katze» oder «Nicht-Katze» gekennzeichnet sind. Das Labeln kann diskriminierend sein, bspw. wenn – sehr vereinfacht

gesagt – ein KI-System geeignete Bewerbende erkennen soll und die als Trainingsdaten verwendeten Dossiers von Personen mit geschützten Merkmalen als «ungeeignet» gelabelt werden.

Werden unausgewogene Daten für das Trainieren von KI-Systemen eingesetzt, führt dies regelmässig zu unausgewogenen und möglicherweise diskriminierenden Ergebnissen («garbage in, garbage out»). Bei selbstlernenden Systemen kann dies zu einer problematischen **Rückkoppelung («feedback loop»)** führen, wenn die Ergebnisse des Systems als Teil einer neuen Datenbasis für das weitere Training des Systems verwendet werden. Ein Beispiel sind Stellenanzeigen in Social Media für eine Lehrperson, die aufgrund stereotypischer Trainingsdaten nur oder überwiegend Frauen angezeigt werden, was zur Folge haben kann, dass vermehrt Frauen diesen Beruf wählen und sich der Frauenanteil im Datensatz laufend weiter erhöht. Ähnliche Effekte können sich einstellen, wenn Menschen diskriminierende Resultate weiterverwenden. Solche Rückkoppelungen können bestehende Diskriminierungen aufrechterhalten, reproduzieren und sogar verstärken.

Auch das **Validieren** und das **Testen** eines trainierten Modells können zu Diskriminierungen führen, wenn die Testdaten unausgewogen sind oder die Qualität des Modells vor allem an den Ergebnissen für gesellschaftliche Mehrheiten (bspw. Menschen mit weisser Hautfarbe) gemessen wird («**Evaluation Bias**»).

Anwendung

Auch bei der Anwendung von KI-Systemen kann es zu Verzerrungen und zu Diskriminierung kommen. Als Beispiele sind zu nennen:

- **Transfer Context Bias:** KI-Systeme können problematisch sein, wenn sie in einem falschen Kontext bzw. in Situationen verwendet werden, für die sie nicht vorgesehen waren. Wenn ein Algorithmus zur Beurteilung einer bestimmten Personengruppe entwickelt wurde, kann er bei der Anwendung auf eine andere Personengruppe zu ungenauen oder falschen Ergebnissen führen
- **Aggregation Bias:** Wenn Rückschlüsse aus einem Einheitsmodell auf Personengruppen gezogen werden, die unterschiedlich betrachtet werden sollten, kann dies zu Diskriminierungen führen. Die Zuordnung von Inputs zu Labels kann zwischen Datensätzen unterschiedlich sein und bestimmte Variablen

können z.B. für Personengruppen mit anderem kulturellem Hintergrund einen anderen Inhalt haben. So können Einheitsmodelle entstehen, die entweder für alle Gruppen nicht richtig passen oder auf die Mehrheit zugeschnitten sind, aber Minderheiten nicht angemessen repräsentieren (bspw. wenn sie auch einen Representation Bias enthalten), wodurch es zu einer diskriminierenden Generalisierung kommen kann.

- **Automation Bias:** Es besteht das Risiko, dass maschinellen Entscheidungen zu viel Vertrauen entgegengebracht wird, weil Menschen die Qualität der Entscheidungen eines KI-Systems überschätzen. So werden Entscheidungen von Systemen, die nur Vorschläge unterbreiten sollen («decision support systems»), in der Realität oft direkt übernommen. Grund dafür ist nicht nur die unterstellte Objektivität und Neutralität des Systems, sondern auch der Zeitdruck bei der Arbeit und der (vermeintliche) Bedarf, eine Abweichung von der Empfehlung zu begründen.

Rechtliche Erfassung

Grundsatz: Technologieneutralität und Erfassen Privater

Die KI-Konvention des Europarates verpflichtet die Schweiz, bestehende Diskriminierungsverbote im nationalen und internationalen Recht auch beim Einsatz von KI zu beachten (Art. 10 (1) KI-Konvention). KI ist jedoch nicht die einzige Technologie, die diskriminieren kann und Diskriminierung kann auch von Menschen ausgehen.

Das geltende Schweizer Recht erfasst in erster Linie Diskriminierung durch staatliche Akteure. Aus Sicht der betroffenen Personen ist es allerdings regelmässig nicht entscheidend, ob sie vom Staat oder von Privaten diskriminiert werden. Entsprechend kennen viele Staaten Regeln, die (auch) Diskriminierung durch Private erfassen. Auch (und gerade) private Unternehmen fällen relevante Entscheidungen zunehmend mithilfe von KI-Systemen (bspw. Gewährung von Krediten oder Selektion von Bewerbungen). Aufgrund der Skalierung kann der Einsatz von KI-Systemen durch Private zudem zu einer Vielzahl diskriminierender Entscheidungen führen und einen gesamtgesellschaftlichen Effekt erzielen, der einer staatlichen Diskriminierung nahekommt. Die Beschränkung von Regeln auf öffentliche Institutionen erscheint deshalb als zu eng.

Herausforderungen

Die Eliminierung von Diskriminierungen in KI-Systemen ist mit Herausforderungen verbunden. Da KI-Systeme über Korrelationen funktionieren, ist das Risiko von Proxy-Diskriminierungen besonders gross. Ob ein geschütztes Merkmal kausal für eine Ungleichbehandlung ist, lässt sich im konkreten Einzelfall nur schwer bestimmen. Dieses Problem wird dadurch verstärkt, dass Daten, die sich auf geschützte Merkmale beziehen, wegen der Vorgaben des Datenschutzrechts oft weder gesammelt noch im System erfasst werden, womit sich Korrelationen kaum erkennen lassen. Versucht man dagegen, Proxy-Diskriminierungen zu verhindern, indem man möglichst wenige Merkmale von Personen in einem KI-System erfasst, entsteht die Gefahr, dass der Algorithmus weniger präzise Resultate generiert.

Allgemeines Gleichbehandlungsgesetz

Der umfassendste Ansatz, um Diskriminierungen durch Private rechtlich zu erfassen, besteht in der Schaffung eines allgemeinen Gleichbehandlungsgesetzes. Ein solches Gesetz könnte auf den bestehenden Gesetzen aufbauen, die schon heute bestimmte Diskriminierungen erfassen, namentlich das Gleichstellungsgesetz (GLG) und das Behindertengleichstellungsgesetz (BehiG). Ein allgemeines Gleichbehandlungsgesetz müsste neben dem verpönten Verhalten auch hinreichend klare Rechtfertigungsgründe und eine Erheblichkeitsschwelle vorsehen, um sachlich gerechtfertigte Ungleichbehandlungen und unerhebliche Fälle auszuschliessen. Zudem wäre denkbar, den Anwendungsbereich eines solchen Gesetzes auf bestimmte Themenfelder zu begrenzen (bspw. Arbeit und Konsum). Neben den materiell-rechtlichen Vorschriften zur Diskriminierung wären auch prozessuale Vorgaben wie eine Beweislastumkehr oder Beweislasteleichterungen und Regeln zu den Verfahrenskosten sowie Sanktionen aufzunehmen.

Ein allgemeines Gleichbehandlungsgesetz hätte den Vorteil, dass es jede Form von Diskriminierung (ev. beschränkt auf bestimmte Themenfelder) vollständig und einheitlich erfassen könnte. Ein neues, umfassendes Gesetz dürfte zudem eine nicht zu unterschätzende Signalwirkung haben und am besten geeignet sein, den Schutz gegen Diskriminierung durch Private sicherzustellen.

Neuer Grundsatz im Datenschutzrecht

Ein zweiter, ebenfalls umfassender Ansatz besteht in der Einführung eines neuen Grundsatzes im Datenschutzgesetz (DSG), nach welchem die Bearbeitung

von Personendaten nicht zu einer Diskriminierung der betroffenen Personen führen darf. Die Grundsätze der Datenbearbeitung statuieren Prinzipien, die bei der Bearbeitung von Personendaten durch den Bund und die Privaten einzuhalten sind, sie gelten für Private aber nicht absolut. Die Nichteinhaltung durch Private kann vielmehr gerechtfertigt werden, namentlich durch Einwilligung der betroffenen Person, durch überwiegende öffentliche oder private Interessen oder durch Gesetz.

Ein neuer Grundsatz im Datenschutzrecht könnte Diskriminierungen weitgehend erfassen, weil diese in aller Regel mit der Bearbeitung von Personendaten verbunden sind. Das gilt auch für Entscheidungen von KI-Systemen, die auf Sachdaten beruhen, weil mit dem Fällen einer Entscheidung für eine bestimmte Person ein Personenbezug hergestellt wird. Anderes gilt nur, wenn ganze Gruppen von Personen diskriminiert werden. Für den Ansatz einer Regelung im Datenschutzrecht spricht, dass er sich durch punktuelle Erweiterung eines bestehenden Gesetzes gesetzgeberisch einfach umsetzen liesse. Hinzu kommt, dass das DSG griffige Durchsetzungsmechanismen enthält und mit dem Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten (EDÖB) eine Behörde besteht, die über die Einhaltung der gesetzlichen Vorgaben wacht und bei Verstössen Verfahren einleiten und Verfügungen erlassen kann.

Die Einführung eines neuen Grundsatzes im DSG erscheint allerdings nicht zwingend. Denn eine zu Diskriminierungen führende Bearbeitung von Personendaten könnte im Rahmen einer teleologischen Auslegung schon unter dem geltenden Recht als Verstoss gegen den datenschutzrechtlichen Grundsatz von Treu und Glauben (Art. 6 Abs. 2 DSG) qualifiziert werden. Für eine solche Auslegung spricht, dass der Zweck des DSG darin besteht, die Grundrechte und die Persönlichkeit der Betroffenen zu schützen (Art. 1 DSG). Die Rechtsgleichheit und das Diskriminierungsverbot (Art. 8 BV) sind grundrechtlicher Natur und Diskriminierungen können als Persönlichkeitsverletzungen qualifiziert werden. Mit Blick auf die Bedeutung des Problems und die Rechtssicherheit wäre es aber sinnvoll, den Grundsatz der Nichtdiskriminierung ausdrücklich im DSG zu normieren.

Diskriminierung als Persönlichkeitsverletzung

Der privatrechtliche Persönlichkeitsschutz (Art. 28 ZGB) greift bei widerrechtlichen Verletzungen der Persönlichkeit. Das Gesetz enthält keine Aufzählung, in

welchen Fällen eine Persönlichkeitsverletzung vorliegt. Lehre und Praxis definieren verschiedene Aspekte der Persönlichkeit, die als solche geschützt sind, wie die physische und psychische Integrität, die Geheim- und Privatsphäre und die Ehre. Die geschützten Aspekte sind nicht abschliessend, sondern unterliegen dem Wandel der Zeit. Werden Entscheide über eine Person getroffen, die auf einem geschützten Merkmal beruhen, welches die Persönlichkeit dieser Person (mit-) definiert, kann eine Persönlichkeitsverletzung vorliegen. Entsprechend spricht sich die Lehre dafür aus, Diskriminierungen als Persönlichkeitsverletzungen zu qualifizieren. Eine Verletzung kann auch hier durch Einwilligung des oder der Verletzten, durch ein überwiegendes privates oder öffentliches Interesse oder durch Gesetz gerechtfertigt werden.

Rechtsdurchsetzung

Ein Schutz gegen Diskriminierungen durch Private wirkt nur, wenn er von den betroffenen Personen (oder Dritten) auch durchgesetzt wird. Den Betroffenen sollten Ansprüche auf Unterlassung und Beseitigung von Diskriminierungen sowie auf Schadenersatz, Gewinnersatz und Genugtuung zustehen. Diese Ansprüche sind im DSG und im allgemeinen Persönlichkeitsrecht schon heute vorgesehen; sie müssten auch in ein allgemeines Gleichbehandlungsgesetz aufgenommen werden. Unabhängig vom materiellrechtlichen Ansatz und den Ansprüchen sind zudem Normen erforderlich, welche den Nachweis von Diskriminierungen und die Durchsetzung der Ansprüche erleichtern.

Der **Nachweis von Diskriminierungen** setzt voraus, dass diese von den betroffenen Personen erkannt und bewiesen werden können. Mit dem Grundsatz der Transparenz, den Informationspflichten und dem Auskunftrecht enthält das DSG zwar schon heute Normen und Mittel, die beim Erkennen und Beweisen von Diskriminierungen durch KI-Systeme eine zentrale Rolle spielen können. Diese Ansprüche sind aber nicht umfassend; sie können namentlich zum Schutz der Geschäftsgeheimnisse von Unternehmen eingeschränkt werden. Das Sammeln von Beweisen ist für die betroffenen Personen zudem oft aufwändig und komplex. Ein griffiger Schutz gegen Diskriminierung dürfte deshalb eine **Beweislastumkehr** oder eine **Beweislast-erleichterung** erfordern, wie sie in ähnlicher Weise bereits im Zusammenhang mit der Prüfung des Diskriminierungsverbots der Bundesverfassung (Art. 8 Abs. 2 BV) erfolgt. Die betroffenen Personen müssten dann

die Anknüpfung an ein geschütztes Merkmal lediglich glaubhaft machen und die beklagte Partei müsste das Vorliegen einer qualifizierten Rechtfertigung beweisen.

Die **Durchsetzung von Ansprüchen** scheitert heute oft daran, dass die von einer Rechtsverletzung betroffenen Personen nicht über ausreichend Zeit und Mittel verfügen, um einen Prozess zu führen. Diese Gefahr besteht auch (und besonders) bei Diskriminierungen. Neben den Ansprüchen der von Diskriminierung betroffenen Personen sollte deshalb ein privatrechtliches Klagerecht für zivilgesellschaftliche Organisationen

eingeführt werden. Der Entwurf zur Änderung der Zivilprozessordnung (ZPO) vom 10. Dezember 2021 sieht hier bereits massgebliche Verbesserungen vor. Namentlich soll die für Persönlichkeitsverletzungen schon bestehende Verbandsklage künftig für alle Rechtsverletzungen zur Verfügung stehen und für Ansprüche auf Schadenersatz und Gewinnherausgabe soll eine neue Verbandsklage geschaffen werden. Diese Mittel des kollektiven Rechtsschutzes erscheinen für das Vorgehen gegen Diskriminierung durch KI-Systeme besonders wichtig, weil hier in der Regel zahlreiche Personen betroffen sind.

Impressum

© 2024

Herausgeberin:
Center for Information Technology,
Society, and Law (ITSL)
Universität Zürich
Rämistrasse 74|38
8001 Zürich