

Woche 14 Recht: Verfassungsrechtliche Basis II – Gleichheitsgrundsatz und
Diskriminierungsverbote mit Blick auf KI

Skript

Erarbeitet von
Jonas Carstens

Lernziele	1
Inhalt	2
Einstieg.....	2
Das Grundgesetz (GG).....	2
Das Allgemeine Gleichbehandlungsgesetz (AGG).....	3
Verschiedene Typen der Diskriminierung durch KI.....	4
Ethische Betrachtung und rechtlicher Rahmen	5
Take-Home Message	6
Quellen	6
Weiterführendes Material.....	7
Disclaimer	7

Lernziele

- Verfassungsrechtliche Basis und Allgemeines Gleichbehandlungsgesetz kennen und unterscheiden können
- Unterscheidung zwischen mittelbarer und unmittelbarer Diskriminierung im Kontext von KI-Anwendungen kennen
- Verständnis für Zusammenhang zwischen KI-Diskriminierung und gesellschaftlicher Ungleichheit stärken

Inhalt

Einstieg

Herzlich willkommen zu AI for All. Diesmal geht es um die rechtlichen Grundlagen der für den KI-Einsatz relevanten Diskriminierungsverbote. Außerdem beschäftigen wir uns erneut mit den ethischen Implikationen der KI-Diskriminierungsproblematik.

Dafür werfen wir zuerst einen Blick auf den Artikel 3 des Grundgesetzes und dann auf das Allgemeine Gleichbehandlungsgesetz. Danach beschäftigen wir uns kurz mit der Frage, wie KI überhaupt diskriminieren kann und schließlich, wie dies mit systemischer Ungleichheit zusammenhängt.

Das Grundgesetz (GG)

Die verfassungsrechtliche Basis des Gleichbehandlungsgrundsatzes bildet der Artikel 3 des Grundgesetzes.

Einblendung des GG Art. 3

Dort heißt es:

- (1) Alle Menschen sind vor dem Gesetz gleich.
- (2) Männer und Frauen sind gleichberechtigt. (...)
- (3) Niemand darf wegen seines Geschlechtes, seiner Abstammung, seiner Rasse, seiner Sprache, seiner Heimat und Herkunft, seines Glaubens, seiner religiösen oder politischen Anschauungen benachteiligt oder bevorzugt werden. Niemand darf wegen seiner Behinderung benachteiligt werden.

Dieser Artikel ist jedoch keineswegs selbsterklärend, wie sich bereits an dem Wort benachteiligen zeigt. So ist zum Beispiel nicht jede formelle Gleichbehandlung automatisch legitim. Auch eine formelle Gleichbehandlung kann bei Betrachtung der Auswirkungen eine Benachteiligung darstellen. Zusätzlich ist zu beachten, dass eine Benachteiligung eine Beeinträchtigung eines rechtlich geschützten Interesses voraussetzt.

Quelle [1]

Außerdem stellt sich die Frage, für wen der Gleichheitsgrundsatz überhaupt gilt und wen er bindet. Als Grundrecht betrifft der Artikel das Handeln des Staates gegenüber Bürger*innen und kann somit auch als Abwehrrecht verstanden werden. Damit gilt der Gleichheitsgrundsatz nicht unmittelbar zwischen Privaten, also beispielsweise nicht

zwischen einem Unternehmen und den Kund*innen. So ergeben sich durch den Artikel andere Anforderungen an ein bei Gericht eingesetztes Modell zur Einschätzung der Rückfallwahrscheinlichkeit als an einen selbstlernenden Credit-Scoring-Algorithmus einer privaten Bank.

Aber nicht jede benachteiligende Ungleichbehandlung ist illegitim. Unter bestimmten Umständen kann diese gerechtfertigt sein. Hier kann zwischen mittelbarer und unmittelbarer Benachteiligung unterschieden werden. Erstere betrifft Regelungen, die sich zwar auf ein neutrales Merkmal beziehen, im Ergebnis aber eine Ungleichbehandlung bedeuten. Unmittelbare Benachteiligung hingegen bezieht sich direkt auf ein geschütztes Merkmal. Mittelbare Benachteiligungen können bereits dann gerechtfertigt sein, wenn „hinreichende sachliche Gründe vorliegen“,

Quelle [2]

während unmittelbare Ungleichbehandlungen nur dann gerechtfertigt sind, wenn sie zwingend erforderlich sind.

Quelle [3]

Schließlich stellt sich die Frage, inwiefern die in Absatz 3 genannte Merkmalsliste als vollständig zu betrachten ist, fehlen doch wichtige Merkmale wie sexuelle Orientierung oder Alter. Die Liste im Artikel 3 ist nicht offen formuliert, sodass ein Hinzufügen eines weiteren Merkmals eine Verfassungsänderung notwendig machen würde und die Liste als abschließend zu betrachten wäre.

Quelle [4]

Bei der sexuellen Orientierung beispielsweise leitet das Bundesverfassungsgericht allerdings ein Quasi-Diskriminierungsverbot her, welches jedoch nicht unumstritten ist, da es richterlich eine Rechtslage herbeiführt, die nicht direkt vom Verfassungsgeber beschlossen wurde.

Quelle [5]

[Das Allgemeine Gleichbehandlungsgesetz \(AGG\)](#)

Jedoch ist der Artikel 3 des Grundgesetzes nicht die einzige gesetzliche Grundlage, die es bei KI und Diskriminierung zu beachten gilt. Von besonderer Bedeutung ist des Weiteren das Allgemeine Gleichbehandlungsgesetz (AGG). Das AGG betrifft auch Private, denn es verbietet Benachteiligungen aufgrund diskriminierungsrelevanter Merkmale, unter anderem bei der Einstellung, bei den Arbeitsbedingungen, bei der Bildung und bei der Versorgung mit Gütern oder Dienstleistungen, die der Öffentlichkeit zur Verfügung stehen. Damit ist das AGG für viele Anwendungskontexte der KI von besonderer Bedeutung. Besonders relevant dürften hier zum Beispiel Kreditverträge

Quelle [6]

und Credit-Scoring-Modelle sein.

Auch im AGG findet sich eine Unterscheidung zwischen mittelbarer und unmittelbarer Benachteiligung. Die Rechtfertigung einer Ungleichbehandlung hängt wieder davon ab, um was für eine Regelung es sich handelt. Beide Formen der Ungleichbehandlung können gerechtfertigt sein. Die Anforderungen an die Rechtfertigung einer unmittelbaren Benachteiligung sind allerdings deutlich strenger. Sie beschränken sich auf bestimmte festgelegte Gründe, wie etwa die Verhütung von Schäden oder den Schutz der Intimsphäre, hier denken wir beispielsweise an die Einstellung ausschließlich weiblicher Psychologinnen in einem Frauenhaus. Mittelbare Benachteiligung kann dann gerechtfertigt sein, wenn sie einem legitimen Ziel dient und angemessen sowie erforderlich ist. Was genau das heißt, werden wir noch an einem Beispiel diskutieren, zunächst widmen wir uns aber noch einmal Fällen der Diskriminierung durch KI.

Verschiedene Typen der Diskriminierung durch KI

Wie genau kann es denn durch den Einsatz von KI zu Diskriminierung kommen? Diese Frage hängt von der spezifischen Anwendung ab und kann pauschal so nicht beantwortet werden. Jedoch lassen sich einige, besonders häufige Probleme identifizieren. Eine umfassende Darstellung findet sich bei Barocas und Selbst in dem Artikel „Big Data’s Disparate Impact“.

Quelle [7]

Ein fast schon klassisches Beispiel betrifft die Diskriminierung durch vorbelastete Trainingsdaten, an denen eine KI trainiert wird. Lernt eine KI zur Bewerber*innenauswahl an einem Set an Trainingsdaten, welches überwiegend männliche Bewerbungen enthält, oder an früheren Personalentscheidungen, die bereits durch Vorurteile geprägt waren, kann die KI lernen, Frauen als weniger geeignet zu behandeln; und zwar unabhängig von den job-relevanten Qualifikationen.

Zusätzlich möchte ich noch auf ein Phänomen aufmerksam machen, das zeigt, wie sehr Diskriminierung mit systemischer Ungleichheit zusammenhängen kann. Hier ist besonders das Phänomen des *redundant encoding* relevant, das zeigt, dass KI auch dann Diskriminierung verstärken kann, wenn keine direkte Diskriminierung durch Trainingsdaten entsteht. *Redundant encoding* beschreibt, dass ein diskriminierungsrelevantes Gruppenmerkmal in einem Teil der Daten zu finden ist, die nicht unmittelbar das Gruppenmerkmal selbst betreffen.

Quelle [8]

Das Beispiel der Kreditrisikobewertung durch KI kann hier Licht ins Dunkel bringen. Dieses Beispiel findet sich besonders eindrücklich in den USA. Trotzdem möchte ich es später anhand des deutschen Rechts diskutieren.

Einblendung Bilder Wohnviertel und Haus aus Geldscheinen

Nehmen wir an, eine KI, eingesetzt bei einer amerikanischen Bank, soll anhand von Daten zu bisherigen Kreditausfällen Korrelationen zwischen bestimmten Merkmalen und der Kreditwürdigkeit von Kandidat*innen finden. Diskriminierungsrelevante Merkmale wie etwa Hautfarbe oder Geschlecht sind bereits aus dem Datensatz entfernt. Nun kann es vorkommen, dass die entdeckten Merkmale sehr eng mit der Mitgliedschaft in einer diskriminierungsrelevanten Gruppe korrelieren. Wenn beispielsweise die Postleitzahl als Prädiktor für Kreditwürdigkeit gewählt wird, dann hätte dieses vermeintlich neutrale Kriterium in den USA den Effekt, dass die als risikoreich erkannten Postleitzahlen zu Bezirken gehören, die durchschnittlich von mehr Afroamerikaner*innen bewohnt werden. Dies hat zwei Gründe: Erstens verfügen Afroamerikaner*innen durchschnittlich über ein geringeres Einkommen und Vermögen. Zweitens haben Jahrzehnte der direkten Diskriminierung bei der Vergabe von Wohnungen und dem Verkauf von Immobilien zu stark segregierten Wohnverhältnissen geführt. So kommt es zu einer schwierigen Situation: Einerseits hätte das Merkmal Postleitzahl für die Bank einen prädiktiven Wert, denn es korreliert eng mit dem Einkommen und dem Vermögen und ist damit von Bedeutung für die Kreditvergabe. Zusätzlich ist es schwierig zu verhindern, dass die KI Merkmale nutzt, die mit diskriminierungsrelevanten Eigenschaften korrelieren. Andererseits bedeutet die Nutzung dieses Merkmals eine weitere Benachteiligung für Afroamerikaner*innen, da das Merkmal der Gruppenzugehörigkeit so eng mit dem Merkmal der Postleitzahl korreliert, dass man von einer Redundanz (*redundant encoding*) sprechen kann. Zusätzlich sind sowohl Vermögens- als auch Wohnverhältnisse durch eindeutig verwerfliche Diskriminierung erst entstanden. Im nächsten Teil widmen wir uns also der Frage, welche Schwierigkeiten solche Phänomene aus ethischer und rechtlicher Perspektive aufweisen.

Ethische Betrachtung und rechtlicher Rahmen

Die beiden vorgestellten Fälle lassen sich anhand einiger wichtiger Merkmale unterscheiden. Erstens liegt im Fall der Trainingsdaten eine unmittelbare Benachteiligung vor, während die Benachteiligung im Fall der Postleitzahl als Kreditvergabekriterium lediglich mittelbar besteht. Außerdem kann im zweiten Fall ein Rechtfertigungsgrund, ein legitimes Interesse der Bank, angeführt werden. Der erste Fall unterläge also auf jeden Fall einem Diskriminierungsverbot durch das AGG. Der zweite Fall stellt sich komplizierter dar. Hier liegt ein legitimes Interesse der Bank vor. Ob dieses Geschäftsinteresse tatsächlich als Rechtfertigung ausreicht, hängt auch davon ab, ob das gewählte Mittel angemessen sowie erforderlich ist, um das entsprechende Ziel zu erreichen. Dahingehend muss insbesondere geprüft werden, ob ähnlich effektive und weniger benachteiligende Merkmale zur Beurteilung der Kreditwürdigkeit herangezogen werden könnten und ob die Verweigerung eines Kredits auf Grund der Postleitzahl einer „übermäßigen Beeinträchtigung“ entspricht.

Quelle [9]

An dieser Stelle hängt die Beurteilung dann wesentlich vom Kontext ab. Grundsätzlich ist aber bei der KI-gestützten Suche nach Korrelationen darauf zu achten, dass bei mittelbaren Benachteiligungen die Angemessenheit überprüft werden sollte.

Take-Home Message

Die ethische Betrachtung zeigt das Problem, was hinter diesen Überlegungen steht, besonders deutlich. Die Nutzung bestimmter Korrelationen aus legitimen Geschäftsinteresse bedeutet die Fortsetzung von Ungleichheiten, die häufig durch Diskriminierung entstanden sind. Solche Probleme der strukturellen Diskriminierung sind aber manchmal nur durch politische Lösungen, in diesem Fall beispielsweise durch staatlich geförderte Kredite, lösbar. Ein einfaches Hinnehmen der Fortsetzung struktureller Diskriminierung erscheint jedenfalls nicht als zufriedenstellende Lösung. Alle, die an KI forschen, diese entwickeln oder einsetzen, sollten sich bewusst sein, dass ein unkritischer Umgang mit dieser Problematik eine Verschärfung gesellschaftlicher Ungleichheiten und Ungerechtigkeiten zur Folge haben kann. Und so ergeben sich sowohl aus rechtlicher als auch aus ethischer Perspektive gute Gründe, mögliche Diskriminierung bei Entwicklung und Einsatz von KI so weit wie möglich zu beachten.

Quellen

- Quelle [1] Jarass/Pieroth/Jarass, 17. Aufl. 2022, GG Art. 3 Rn. 10-15
Jarass, H.D. & Pieroth, B. (2022). *Grundgesetz für die Bundesrepublik Deutschland: GG Kommentar* (17. Auflage). C.H. Beck, München.
- Quelle [2] BeckOK GG/Kischel, 44. Aufl. 15.8.2020, GG Art. 3 Rn. 190-190a
Epping V. & Hillgruber C. (2020). *Beck'scher Online-Kommentar GG* (44. Auflage).
- Quelle [3] BeckOK GG/Kischel, 44. Aufl. 15.8.2020, GG Art. 3 Rn. 191-195.1, 191
- Quelle [4] BeckOK GG/Kischel, 44. Aufl. 15.8.2020, GG Art. 3 Rn. 231
- Quelle [5] BeckOK GG/Kischel, 44. Aufl. 15.8.2020, GG Art. 3 Rn. 130-131.1
- Quelle [6] Schleusener /Suckow/Plum, AGG - Kommentar, 6. Auflage 2022 RN17
Schleusener, A., Suckow, J. & Plum, M. (2022). *AGG – Kommentar* (6. Auflage). C.H. Beck, München.
- Quelle [7] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review* 104, S. 671-732.
- Quelle [8] Barocas/Selbst 2016, S. 681
- Quelle [9] Bundesarbeitsgericht (BAG), Urteil vom 15.12.2016 - Aktenzeichen 8 AZR 454/15

Weiterführendes Material

Verschiedene philosophische Theorien zur Unterscheidung zwischen legitimer und verwerflicher Diskriminierung:

Helmann, D. (2008). *When Is Discrimination Wrong?* Harvard University Press, Cambridge.

Lippert-Rasmussen, K. (2006). The badness of discrimination. *Ethical Theory and Moral Practice* 9(2). S. 167-185.

Sangiovanni, Andrea (2017). *Humanity without Dignity*. Harvard University press, Cambridge.

Wie neue technologische Entwicklungen, insbesondere Algorithmen und KI, Diskriminierung und Rassismus verstärken (gut verständlich):

Benjamin, R. (2019). *Race after Technology*. Polity Press, New York.

Eubanks, Virginia (2019). *Automating Inequality How High-Tech Tools Profile Police, and Punish the Poor*. Picador, New York.

Kritik an technozentrischer Perspektive und Ruf nach mehr Einbindung gesellschaftlicher Strukturen in der KI-Entwicklung:

Bui, M. L., & Noble, S. U. (2020). We're missing a moral framework of justice in AI. In Duber, M. D., Pasquale, F., & Das, S. (Hrsg.) *The Oxford Handbook of Ethics of AI*. (S.163-181), Oxford University Press, Oxford.

Selbst, A. D., Danah, B., Friedler, S. A., Venkatasubramanian S., & Vertesi, J. (2018). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, S. 1-17.

Disclaimer

Transkript zu dem Video „Woche 14 Recht: Verfassungsrechtliche Basis II – Gleichheitsgrundsatz und Diskriminierungsverbote mit Blick auf KI“, Jonas Carstens.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.