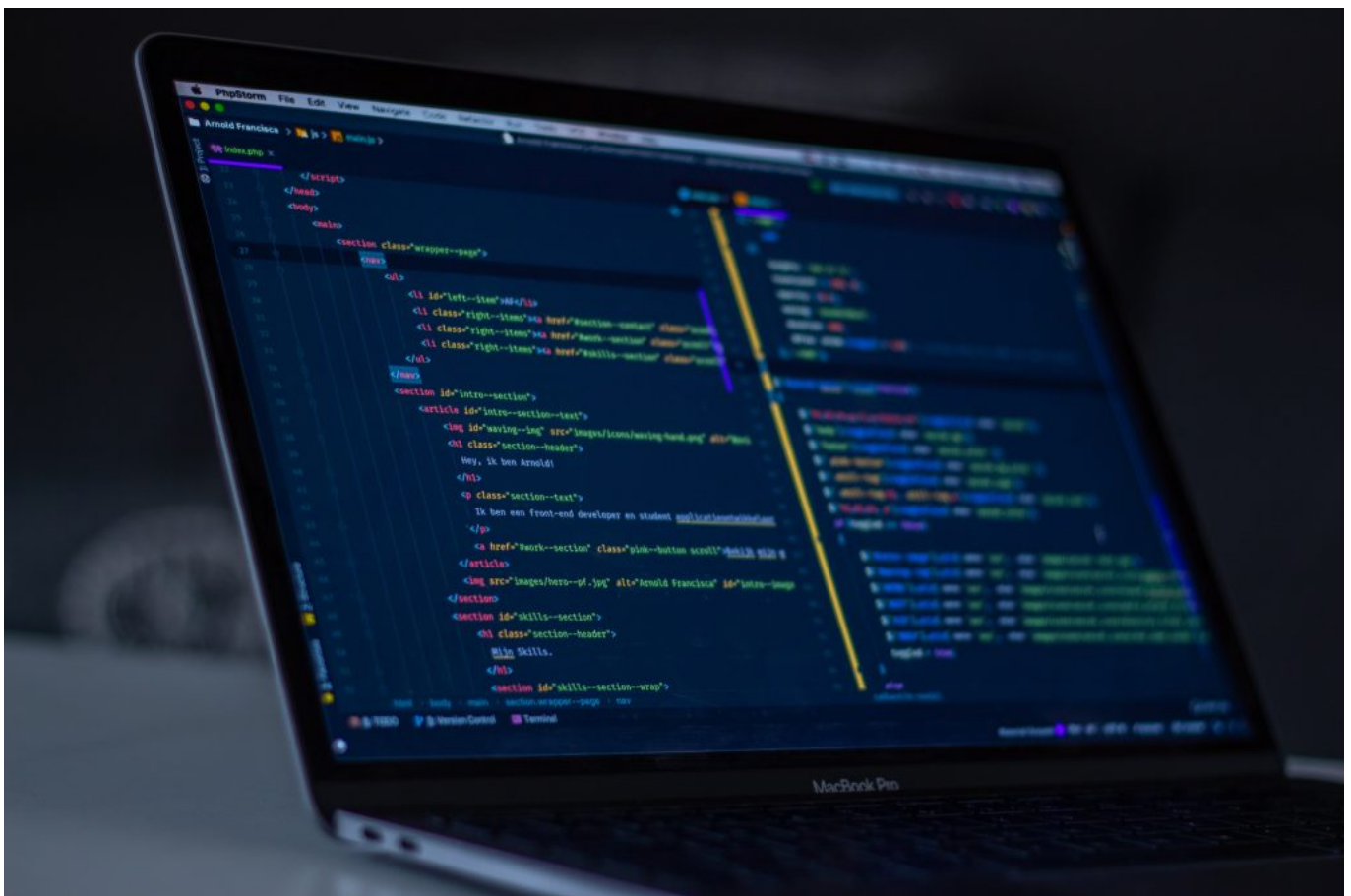


ALGORITHMISCHE DISKRIMINIERUNG: “ALGORITHMEN ALS EIN TEIL VON JENER KRAFT, DIE STETS DAS GUTE WILL UND STETS DAS BÖSE SCHAFFT.”

Fabian Lütz
18th Juli 2023



Algorithmen und die zugrunde liegenden Datensätze sind nicht neutral, sondern spiegeln vorherrschende Vorurteile und Biases der Gesellschaft. Algorithmen können bei der Erreichung der Gleichstellungsziele hinderlich sein, weil sie insbesondere aufgrund der Phänomene des Gender Bias, Gender Data Gap und des Machine Bias diskriminierende Wirkungen entfalten können. Algorithmen können aber auch gezielt für die Erreichung von Gleichstellungszielen eingesetzt werden, unter anderem zur Verfolgung positiver Massnahmen und zur Aufdeckung von Diskriminierungen. Am Beispiel von Rekrutierungsalgorithmen lassen sich die negativen und positiven Konsequenzen des Einsatzes von Algorithmen für die Gleichstellung von Männern und Frauen anschaulich erklären.

Erst kürzlich wurde im Schweizer Parlament eine Interpellation im Hinblick

auf die Regulierung von Künstlicher Intelligenz hinterlegt.¹ In Europa und auch in der Schweiz gibt es Vorschriften, die eine Diskriminierung aufgrund des Geschlechts verhindern sollen. Durch den vermehrten Einsatz von Künstlicher Intelligenz und Algorithmen kann es zu sogenannten algorithmischen Diskriminierungen kommen. Unter algorithmischer Diskriminierung versteht man eine ungerechtfertigte Ungleichbehandlung aufgrund eines geschützten Merkmals (z.B. Geschlecht oder Religion), bei der eine menschliche Entscheidung ganz oder teilweise durch einen Algorithmus ersetzt oder wesentlich unterstützt wird und die Diskriminierung eines Menschen bewirkt.

Ursache hierfür können entweder das Design der Algorithmen selbst oder aber die Datensätze sein, welche Vorurteile oder Biases enthalten. Die Datensätze werden sowohl für das Training, die Validierung, als auch zum Testen der Algorithmen benötigt. Ein konkretes Beispiel von algorithmischer Diskriminierung ist der Einsatz von Rekrutierungssoftware, bei der Algorithmen die Personalauswahl unterstützen oder sogar ersetzen. Eine Diskriminierung ist hier auf allen Stufen des Bewerbungsverfahrens möglich, Einfallstor ist jedoch häufig das automatische Sichten von Lebensläufen (sog. CV parsing²). Hier kann einerseits das Design des Algorithmus durch Vorurteile und Biases der EntwicklerInnen beeinflusst worden sein, weil das *mindset* zu gewissen Entscheidungen beim Entwerfen eines Modells für den Algorithmus führt. Andererseits können die für das Trainieren des Modells verwendeten Datensätze vorurteilsbehaftet sein, weil es sich notwendigerweise um historische Daten handelt. So können beispielsweise im IT-Bereich die Daten von früheren Bewerbungen zum Trainieren des Modells verwendet werden, um die idealen Job-Kandidaten zu gewinnen. Wenn ausschliesslich oder vermehrt männliche Kandidaten in der IT-Branche arbeiteten oder arbeiten, dann wirkt sich das auf die Entscheidungen des Algorithmus aus, weil dieser auf der Basis der verfügbaren Daten entscheidet. Hier beeinflussen beispielsweise der Gender Data Gap und der Digital Gender Divide die vorhandenen Daten und somit die Ergebnisse der Algorithmen-basierten Entscheidungen im Rekrutierungsbereich.³

Grundsätzlich gilt, dass der gleiche Schutz vor Diskriminierung sowohl offline als auch in der online-Welt gewährleistet sein sollte. Viele Beispiele aus jüngster Zeit zeigen jedoch, dass Algorithmen Biases oder sogar Diskriminierungen hervorrufen können.⁴

Aus diesen Gründen haben seit einigen Jahren internationale Organisationen (z.B. Europarat⁵, OECD⁶, UNESCO⁷) und Staaten (z.B. Kanada⁸, USA⁹, Schweiz¹⁰) Empfehlungen im Bereich der Menschenrechte und künstlicher Intelligenz verabschiedet, welche in nicht bindender Weise Handlungsvorschläge zum Umgang mit künstlicher Intelligenz geben. Insbesondere die Risiken von KI im Hinblick auf Gleichstellung und Diskriminierung werden in den jeweiligen Dokumenten behandelt. Aufgrund verstärkter Diskussionen in der Öffentlichkeit und der Wissenschaft, diskutiert die Politik mittlerweile auch über rechtliche Rahmen, um den Einsatz von Künstlicher Intelligenz zu regulieren, hauptsächlich um negative Konsequenzen wie beispielsweise Biases und Diskriminierungen zu verhindern.

SCHWEIZ UND DER EUROPARAT

In der Schweiz gibt es derzeit keine konkreten juristischen Vorstösse, um künstliche Intelligenz zu regulieren. Allerdings gibt es einige nichtbindende Dokumente, die insbesondere Leitlinien für künstliche Intelligenz und den öffentlichen Dienst skizzieren. Die Schweiz ist allerdings Mitglied des Europarats, der im Jahre 2020 eine Empfehlung zu Künstlicher Intelligenz und Menschenrechten verabschiedet hat. Der Europarat arbeitet momentan an einer Rahmenrechts-Konvention, um die menschenrechtlichen Auswirkungen beim Einsatz von Künstlicher Intelligenz zu regeln. Das zuständige Komitee, das *Committee on Artificial Intelligence* (CAI), zurzeit unter dem Vorsitz des Schweizer Thomas Schneider, ist für die Ausarbeitung des rechtlichen Vorschlags verantwortlich.¹¹ Sollte der Vorschlag Ende des Jahres 2023 angenommen werden, so könnten nach Annahme und Ratifikation durch die Schweiz diese Regeln auch in der Schweiz geltendes Recht werden.

EUROPÄISCHE UNION – VORREITERROLLE BEI DER REGULIERUNG VON KI

Darüber hinaus hat die Europäische Union einen Gesetzgebungsvorschlag zu künstlicher Intelligenz im Jahr 2021 vorgestellt, den sogenannten EU Artificial Intelligence Act („EU AI Act“)¹², welcher momentan von den europäischen Gesetzgebern im sogenannten Trilog Verfahren, unter spanischer Ratspräsidentschaft, verhandelt wird. Sollte der Vorschlag angenommen werden, würde es sich um die ersten rechtlichen Regeln weltweit handeln, die konkrete Vorschriften beim Einsatz von Künstlicher Intelligenz insbesondere für Unternehmen vorsehen. Zwar handelt es sich bei dem Vorschlag für eine Verordnung nicht um eine spezifische Regelung, um Gleichstellung zu erzielen und Diskriminierungen zu bekämpfen, jedoch fallen bestimmte Hoch-Risiko KI-Systeme in den Anwendungsbereich, die sehr relevant und möglicherweise das Risiko einer Diskriminierung mitbringen. Ein Beispiel sind die sogenannten KI-Rekrutierungssysteme, welche vom Anwendungsbereich der zukünftigen Verordnung umfasst sind, und welche in der Zukunft bestimmten Regulierungsanforderungen unterliegen werden, beispielsweise im Hinblick auf Transparenz, Dokumentation oder Zugang zum Quellcode der Algorithmen etc. Diese Regulierungsanforderungen könnten es potenziellen Opfern einer Diskriminierung erleichtern, eine Beschwerde oder Klage erfolgreich durchzuführen.

VEREINTE NATIONEN – GLOBALE REGELN FÜR DAS GLOBALE PHÄNOMEN KI

Auch die Vereinten Nationen denken über Möglichkeiten nach, künstliche Intelligenz zu regulieren, um mögliche Menschenrechtsverletzungen zu vermeiden. So wurden beispielsweise einige Berichte von Sonderberichterstatern verabschiedet, die Aspekte von künstlicher Intelligenz, Algorithmen und den Auswirkungen auf Diskriminierung und Gleichstellung untersuchen und konkrete Handlungsvorschläge und Empfehlungen sowohl an Staaten als auch an Unternehmen beinhalten.¹³ Die Gruppe *The Elders* - ehemalige Staatsmänner und -frauen, Friedensaktivist:innen, Menschenrechtler:innen sowie prominente Intellektuelle - riefen kürzlich die Vereinten Nationen (VN) auf, eine neue globale Regulierungsarchitektur für Künstliche Intelligenz zu schaffen sowie einen völkerrechtlichen Vertrag auszuarbeiten, der eine internationale Agentur für die Sicherheit der

Künstlichen Intelligenz schaffen soll. ¹⁴

TECHNOLOGIEUNTERNEHMEN – UNVERBINDLICHE KI-PRINZIPIEN UND ETHISCHE RICHTLINIEN

Schliesslich haben die meisten Technologieunternehmen auch selbst nicht bindende Prinzipien für künstliche Intelligenz oder ethische Empfehlungen verabschiedet, die viele Aspekte wie Diskriminierung, Gleichstellung, Vorurteile, Diversität und Ähnliches beinhalten.¹⁵ Allerdings fehlen diesen Empfehlungen nicht nur die bindende Wirkung, sondern auch konkrete Durchsetzungsmöglichkeiten, falls Bürger:innen Opfer einer Diskriminierung werden.

AUSBLICK

Es kann mit Spannung verfolgt werden, ob die Schweiz unabhängig von den Vorschlägen der Europäischen Union und des Europarats eigene Vorschläge auf den Tisch legt. Zuletzt gab es einige Versuche, das Thema im Schweizer Parlament auf die Agenda zu setzen.¹⁶ Die möglichen negativen Auswirkungen der KI auf Gleichstellung und Diskriminierung, aber auch die potenziellen Chancen, um negative Folgen zu erkennen und zu bekämpfen, machen eine Beschäftigung mit dem Thema sowohl in akademischen als auch in politischen Zirkeln unerlässlich. Die Öffentlichkeit ist spätestens seit *ChatGPT* für die Chancen und Gefahren der KI sensibilisiert und es bleibt mit Spannung abzuwarten, ob die Schweiz auf den Regulierungszug aufspringt. Auch wenn Pünktlichkeit als eine Tugend der Schweiz gilt, werden die rechtlichen Regeln im Bereich der KI wohl noch ein bisschen auf sich warten lassen.

Quellen:

- Lütz, Fabian. (2023) Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung, GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft, 1-2023, S. 26-41.
<https://doi.org/10.3224/gender.v15i1.03> .
- Lütz, Fabian. (2023). Gender Equality and Artificial Intelligence: SDG 5 and the Role of the UN in Fighting Stereotypes, Biases, and Gender Discrimination. In: Fornalé, E., Cristani, F. (eds) Women's Empowerment and Its Limits. Palgrave Macmillan, Cham.
https://doi.org/10.1007/978-3-031-29332-0_9 .
- Lütz, Fabian. (2023). Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé, dans Guillaume Florence (eds.) La technologie, l'humain et le droit (Conférences CUSO), Stämpfli Verlag.
- Lütz, F. Gender equality and artificial intelligence in Europe. Addressing direct and indirect impacts of algorithms on gender-based discrimination. ERA Forum 23, 33–52 (2022).
<https://doi.org/10.1007/s12027-022-00709-6> .

Referenzen:

Titel: Johann Wolfgang von Goethe: Faust 1 - Hamburger Ausgabe Band 3, dtv, München 1982, S. 47, Studierzimmer, 1334-1336; Im Original von Johann Wolfgang von Goethe heisst es: "Ich bin ein Teil von jener Kraft, Die stets das Böse will und stets das Gute schafft."

[1] Die Bundesversammlung, 23.3147, INTERPELLATION vom 14. März 2023, Regulierung der künstlichen Intelligenz in der Schweiz, <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20233147> .

[2] Vgl. [https://www.talention.de/blog/was-ist-cv-parsing-eine-erklaerung-und-die-vorteile-auf-einen-blick#:~:text=Resume%20Parsing%20\(CV%20%3D%20Curriculum%20vitae,in%20eine%20Bewerbermanagement%2DSoftware%20importiert.](https://www.talention.de/blog/was-ist-cv-parsing-eine-erklaerung-und-die-vorteile-auf-einen-blick#:~:text=Resume%20Parsing%20(CV%20%3D%20Curriculum%20vitae,in%20eine%20Bewerbermanagement%2DSoftware%20importiert.)

[3] Vgl. weiterführend zu Diskriminierung bei Rekrutierungsalgorithmen, Fabian Lütz, Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé, dans Guillaume Florence (eds.) La technologie, l'humain et le droit (Conférences CUSO), Stämpfli Verlag (2023).

[4] Dastin, Jeffrey. 2022. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> (Rekrutierungen-Algorithmen); Vigdor, Neil. 2019. Apple Credit Card Investigation. The New York Times, November 10 (verschiedene Kreditrahmen für Männer und Frauen).

[5] Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154 .

[6] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> .

[7] UNESCO, Recommendation on the Ethics of Artificial Intelligence, <https://unesdoc.unesco.org/ark:/48223/pf0000381137> .

[8] Canada, Responsible use of artificial intelligence (AI), <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1> .

[9] US White House, Blueprint for an AI Bill of Rights, MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> .

[10] Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung (2020), https://www.sbf.admin.ch/dam/sbf/de/dokumente/2020/11/leitlinie_ki.pdf.download

[load.pdf/Leitlinien%20Künstliche%20Intelligenz%20-%20DE.pdf](#) .

[11] <https://www.coe.int/en/web/artificial-intelligence/cai>

[12] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM/2021/206 final.

[13] Vergleiche z.B. A/HRC/44/57, Report of the Special Procedure of the Human Rights Council. Racial discrimination and emerging digital technologies: Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance; A/HRC/48/31, The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights; A/HRC/53/65, Digital Innovation, technologies and the right to health, Report of the Special Rapporteur (21 April 2023); A/HRC/53/24, Building capacity for the implementation of the Guiding Principles on Business and Human Rights, Report of the working group on the issue of human rights and transnational corporations and other business enterprises; A/HRC/53/L.27/Rev.1, New and emerging digital technologies and human rights (highlighting both the risks for the right to equality and non-discrimination and the opportunities for gender equality.

[14]

<https://theelders.org/news/elders-urge-global-co-operation-manage-risks-and-share-benefits-ai>

[15] Vgl. zum Beispiel, Google (AI Principles, available at: <https://ai.google/principles/>); Amazon (Responsible use of artificial intelligence and machine learning), available at: <https://aws.amazon.com/de/machine-learning/responsible-machine-learning/>), Meta (Facebook's five pillars of Responsible AI, available at: <https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/>) ; Microsoft (Microsoft Responsible AI Standard, v2, June 2022, available at: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl>) .

[16]

<https://www.tdg.ch/le-ps-passe-a-loffensive-sur-lintelligence-artificielle-807307585264> ; <https://www.parlament.ch/fr/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20233147> .

Bild: [unsplash.com](#)