­GGW MZ 2018 1

ca 3200 Zeichen

Ziel: maximal 7.500 Zeihen

**Überschrift**

*Gert G. Wagner und Meike Zehlike* ***[wenn Sie mehr beitragen als ich auch gerne gegen das Alphabet: MZ und GGW]***

*Der Berliner Ökonom Gert G. Wagner ist Mitglied der Deutschen Akademie der Technikwissenschaften (acatech) und Mitglied im Sachverständigenrat für Verbraucherfragen. Die Informatikerin (?) Meike Zehlike ist xxxx Technischen Universität Berlin.*

Im Aufkommen des maschinellen Lernens und damit der Möglichkeit, automatisiert Entscheidungen zu treffen, herrscht unter Technologen zunächst große Euphorie. Endlich würden Entscheidungen rational und effizient getroffen werden können, ohne dass das Tagesbefinden oder die emotionale Prägung eines vormals menschlichen Entscheiders eine Rolle spielen würde. „Algorithmen entscheiden ausschließlich auf Basis von meritokratischen Fakten, nicht auf Basis von Hautfarbe“ und „Mathematik und Algorithmen sind neutral und deswegen sind es auch algorithmische Entscheidungen“ sind gängige Leitsätze dieser Zeit, die noch nur langsam überwunden wird, so scheint es. Im Rahmen einer Markterhebung, die der Sachverständigenrat für Verbraucherfragen jüngst durchführte, schrieb eine Auskunftei: „Mathematisch-statistische Verfahren diskriminieren nicht.“

Es stellt sich nun allerdings heraus, dass diese Aussagen irreführend und z.T. schlicht falsch sind. Algorithmen sind weder per se neutral, noch ist es unmöglich, dass selbstlernende Systeme diskriminieren, ganz im Gegenteil. Gerade bei selbstlernenden Systemen ist die Gefahr einer systematischen Diskriminierung durch das erlernte Modell groß, denn diese werden auf historischen Daten trainiert, die als Abbild der damaligen und heutigen Gesellschaft sämtliche historischen und aktuellen Diskriminierungen enthalten. Jüngst erregte der Konzern Amazon mit einem System Aufsehen, dass automatisch ungeeignete Bewerbungen auf Programmierjobs aussortieren sollte. Da Frauen\* in der Informatik stark unterrepräsentiert sind und damit ein statistisches Handycap aufweisen, lernte das System, dass Bewerbungen von Frauen\* generell schlechter zu bewerten sind. Dies ist in zweierlei Hinsicht problematisch: zunächst sind nun *alle* Bewerbungen von Frauen\* von einem sexistischen Algorithmus betroffen und nicht nur die Unglückseeligen, die das Pech hatten an einen sexistischen Bearbeiter zu geraten. Zusätzlich besteht nun die Schwierigkeit, dass ein Opfer solcher Diskriminierung es sehr schwer hat, einerseits seinen Opferstatus überhaupt zu erkennen, andererseits dagegen vorzugehen. Es diskutiert sich nämlich schlecht mit Algorithmen. Die Welt ist also darauf angewiesen, dass diejenigen, die ein solches System einsetzen, ihre Hausaufgaben gemacht haben.

**Frage: Wie kommen wir von hier dahin, dass Algorithmen trotzdem eine gute Alternative sein können?**

**Einige Argumente:**

* **Algorithmen sind effizienter und können getrost in Bereichen eingesetzt werden, wo eine Entscheidung egal wie sie ausfällt keine negativen Konsequenzen nach sich zieht (trifft aber nicht auf das Amazon-Beispiel zu)**
* **Mit einem geeigneten Testset lässt sich durchaus feststellen, ob ein Algorithmus diskriminiert (hätte Amazon in einer Testphase darauf geachtet, dass deutlich weniger Frauen eingeladen werden, als sich bewerben, hätte man die Diskriminierung nicht nur feststellen, sondern sogar quantifizieren können) → das geht allerdings auch, wenn Menschen solche Entscheidungen treffen, hier sehe ich noch nicht, wieso da ein Algorithmus einen Vorteil brächte.**
* **Wenn mensch es einmal geschafft hat, einen Algorithmus zu bauen, der nicht diskriminiert, dann wären wir alle Sorgen auf einmal los. Man müsste keine Ressourcen mehr in Mitarbeiterawareness stecken und alle würden auf einen Schlag davon profitieren…**
* **Man kann verschiedene Fairness-Modelle bauen, die hinsichtlich verschiedener Kriterien optimieren**

Künstliche Intelligenz und „Algorithmen“ werden in der Öffentlichkeit und etwa in geisteswissenschaftlichen Seminaren skeptisch betrachtet, weil man unter anderem Angst davor hat, dass Algorithmen Menschen unfair oder gar diskriminierend behandeln. Wenn etwa Frauen xxxxxx und Angehörige bestimmter Ethnien yyyyyy. Das kann in der Tat passieren, wenn man Algorithmen naiv einsetzt – wie man es bei Informatikern und Firmen, die die Bonität von uns prüfen, vielfach beobachten kann. Im Rahmen einer Markterhebung, die der Sachverständigenrat für Verbraucherfragen jüngst durchführte, schrieb eine Auskunftei: „Mathematisch-statistische Verfahren diskriminieren nicht.“ ***[MZ: siehe Seite 86 des Gutachtens]*** Wenn man Algorithmen auf Basisw einer solchen Annahme anwendet, können sie in der Tat schwer in die Irre führen. Wir wollen hier aber darauf hinweisen, dass algorithmische Verfahren helfen können Unfairness und Diskriminierung offen zu legen, die bei menschlichen Entscheidern schwer nachzuweisen wären. Und in der Tat entwickelt sich gerade eine xxxxxx Forschung.

Zuerst einmal muss festgehalten werden, dass es sehr schwer sein kann einem menschlichen Entscheider Unfairness oder Diskriminierung nachzuweisen. Wenn ein Privatmann eine Wohnung vermietet wird man ihm niemals Diskriminierung nachweisen können, wenn er nicht ausspricht, dass er etwa Menschen mit dunkler Hautfarbe oder Familien mit kleinen Kindern nicht mag. Würde ein Computeralgorithmus diesem Vermieter helfen Bewerber vor zu sortieren, dann könnte man Diskriminierung ohne weiteres feststellen, wenn sie einprogrammiert ist. Dies kann man im übrigen auch einem „neuronalen Netz“, also „künstlicher Intelligenz“, nachweisen, von der zwar gerne behauptet wird, dass sie „selbst lerne“ und undurchschaubar wäre. Das stimmt aber schlicht und einfach nicht: auch künstliche Intelligenz kann man testen, so wie die Stiftung Warentest Kaffeemaschinen und Computer testet.

Also: Algorithmen bieten ohne weiteres die Chance Unfairness und Diskriminierung zu erkennen. Und mit Hilfe von Algorithmen kann man sogar besser erkennen und öffentlich diskutieren, was Unfairness und Diskriminierung eigentlich ist und wie man ihnen begegnen kann. Insbesondere macht die Fairness-Forschung mit Hilfe von Algorithmen deutlich wie viele Spielarten von Fairness gibt – und dass in der Regel man nicht alle Dimensionen gleichzeitig erreichen kann.

XXXXX

MÖGLICHER SCHLUSS Bislang kennt sich die einschlägige Community der Forscher zur Künstlichen Intelligenz wenig mit der Forschung zu Fairness aus ethischer und philosophischer Sicht aus und es fällt deswegen schwer, die verschiedenen Ansätze so einzuordnen, dass der Laie – also wir alle – wissen kann, in welcher Situation er welchen Algorithmus anwenden sollte. Deswegen ist es notwendig, dass KI-Forschung genauer hinsieht, was Fairness eigentlich in welcher Situation bedeutet. Ansonsten besteht die Gefahr, dass Scoring-Unternehmen sich einfach irgendeine Variante auf gut dünken aussuchen und sich dahin dahinter verstecken, dass ihre Software fair sei obwohl sie potentiell die gleichen Probleme wie vorher haben oder sie sogar noch verschlimmern, weil denen gar nicht klar ist, „welche“ Fairness sie eigentlich optimieren.