

Nachdem Sie endlich die Ressourcen für Ihre Arbeit erhalten haben, Ihre Arbeit getan ist, die Filter hoffentlich vernünftig funktionieren, Malware erkannt wird, die Radio-Experten das Netzwerk am Laufen halten können und den Gesetzen genüge getan ist, trinken Sie erst mal einen großen Kaffee und grübeln darüber nach, ob die Sicherheit und Privacy der Benutzer jetzt wohl wirklich zu 100 % geschützt sind und ob wohl alles der *Privacy Policy* entspricht? Als Ihr Kaffeebecher leer ist, sieht er so aus, als hätte eine Pille auf dem Grund gelassen. Er ist manchmal rot, manchmal blau ...

erschienen in der *FIfF-Kommunikation*,
herausgegeben von *FIfF e.V.* - ISSN 0938-3476
www.fiff.de

Nachtrag

Die Autoren haben diese Geschichte natürlich komplett erfunden. Die Geschichte und die hier repräsentierten Meinungen sind die der Autoren und nicht notwendigerweise die von *Nokia Networks*. Und unser Manager hat ein gutes Sicherheitsver-

ständnis. Wir haben einige technische Details verallgemeinert, um die Leser nicht mit 3-5-Buchstaben-Abkürzungen zu bombardieren. Falls aber Interesse an selbigen besteht, werden wir diese selbstverständlich auf direkte Anfragen genau erläutern. Das Thema Sicherheit der Luftschnittstellen haben wir bewusst nicht genauer erläutert, da es recht heterogen ist. In diesem Bereich gibt es viele Varianten die den Rahmen des vorliegenden Textes sprengen würden und einen eigenen Artikel verdienen.

- 1 Wenn Benutzer HTTPS Ende-zu-Ende benutzen, ist das keine Maßnahme, die vom Netzwerkbetreiber angewandt wird
- 2 Status 2012, es geht nicht daraus hervor, ob diese Zahlen auch die Virtual Operator enthalten
- 3 Chaos Computer Club, Dez. 2014, Tobias Engel



Max Maaß

Faire Algorithmen

In der öffentlichen Diskussion werden Algorithmen oft als objektiv oder gar unfehlbar dargestellt. Ein Algorithmus habe schließlich keine versteckten Vorurteile, er behandle alle gleich, und sei damit per definitionem objektiv. Diese Einschätzung nutzen einige als Argument, um immer mehr Aufgaben von Menschen auf Algorithmen zu übertragen – von der Verkehrsplanung über die Auswahl von Bewerber:innen bis hin zur Strafverfolgung mittels Predictive Policing.¹ In diesem Artikel soll die Frage untersucht werden, ob Algorithmen tatsächlich objektiv sind, und welche Probleme sich ergeben können, wenn Algorithmen die Arbeit von Menschen übernehmen.

Eine Geschichte unfairer Algorithmen

Die Untersuchung von Algorithmen auf Verfälschungen und Befangenheit² geht bereits mehr als 30 Jahre zurück. Schon in den 1980ern wurde dem Buchungssystem *Sabre* (betrieben von *American Airlines*) vorgeworfen, bestimmte Flüge bevorzugt anzuzeigen. Tatsächlich stellte sich heraus, dass das System von *American Airlines* gezielt dazu verwendet wurde, konkurrierende Fluglinien systematisch zu benachteiligen.³ Dadurch erhielt *American Airlines* einen Vorteil gegenüber der Konkurrenz, der sich auch in den Bilanzen niederschlug.

In diesem Fall wurde der Algorithmus gezielt so entworfen, dass er bestimmte Ergebnisse bevorzugt. Schwieriger wird es beim anderen Extrem: Ein Algorithmus, der möglichst neutral sein soll, aber *trotzdem* befangen ist. Ein Beispiel dafür war ein Algorithmus, der in den 1970ern von einer englischen *Medical School* entwickelt wurde, um den Bewerbungsprozess um die limitierten Studienplätze zu beschleunigen. Das Programm sollte eine erste Auswahl der Kandidat:innen vornehmen, die für ein Bewerbungsgespräch eingeladen werden sollten. Es wurde über mehrere Jahre weiterentwickelt, bis es eine Übereinstimmung von bis zu 95 % mit den Ergebnissen der (menschlichen) Bewerbungskommission ergab.

Der Algorithmus war über fünf Jahre im Einsatz, bis sich herausstellte, dass er gegen Frauen und ethnische Minderheiten diskriminierte. Dieses Verhalten wurde ihm nicht explizit beigebracht,

er hatte es aus den Entscheidungen der Kommission gelernt, die (bewusst oder unbewusst) eine solche Diskriminierung an den Tag gelegt hatte.⁴

Der Stand der Forschung

Diese Beispiele zeigen, dass Algorithmen sowohl gezielt als auch unbeabsichtigt verzerrend wirken können. Während *analoge* Diskriminierung schon seit längerer Zeit wissenschaftlich untersucht wird, ist das Feld der Erforschung digitaler Diskriminierung vergleichsweise wenig erschlossen. Die erste Arbeit zu diesem Thema wurde 1987 von Huff *et al.* geschrieben und weist auf die Gefahr von Gender-Stereotypen in der Softwareentwicklung hin.⁵ Diese Arbeit stammt allerdings aus der Psychologie, nicht der Informatik.

Die ersten Informatikerinnen, die sich mit diesem Themenkomplex beschäftigten, waren Batya Friedman und Helen Nissenbaum, die 1996 einen Artikel über die verschiedenen Formen von Verfälschungen in Computersystemen schrieben. Sie identifizierten drei Formen von Verfälschung: *preexisting bias* (Tendenzen, die aus der realen Welt übernommen werden), *technical bias* (Verfälschungen durch technische Limitationen), und *emergent bias* (Verfälschungen, die erst im Verlauf der Zeit entstehen, weil sich die Software nicht an neue Begebenheiten anpasst).⁶ Der bereits diskutierte Fall des Bewerbungs-Algorithmus wäre in diesem Modell ein Fall von *preexisting bias*. Ein Beispiel

für *technical bias* wäre, gleichwertige Ergebnisse alphabetisch zu ordnen, wodurch einige Ergebnisse durch prominentere Positionen in der Ergebnisliste systematisch bevorzugt werden. Ein häufiger Fall von *emergent bias* sind binäre Auswahlfelder für Geschlechter, die keine abweichenden Geschlechtsidentitäten zulassen.

Über die nächsten Jahre beschäftigen sich auch andere Wissenschaftler:innen mit dem Thema. 2005 studiert Phoebe Sengers, wie unterbewusste kulturelle Annahmen das Softwaredesign beeinflussen können.⁷ Einige Arbeiten beschäftigten sich mit den Algorithmen von Werbenetzwerken, die teilweise gegen Frauen und ethnische Minderheiten diskriminierten.⁸ Im Jahr 2014 schrieb Nicolas Diakopoulos einen Report über *algorithmic accountability reporting*,⁹ also das Untersuchen und kritische Hinterfragen von Algorithmen (auch) durch Journalisten. Im selben Jahr schreibt Tal Z. Zarski einen Artikel über *discrimination in the scored society*,¹⁰ in dem er Formen der Diskriminierung in *Scoring*-Verfahren (wie der Berechnung der Kreditwürdigkeit) analysiert.

Algorithmische Kollateralschäden

Der Schaden, den unfaire Algorithmen anrichten, bleibt für Nicht-Betroffene meist unsichtbar. Er trifft häufig die, die bereits benachteiligt sind, etwa durch die Verfestigung von Geschlechter-Stereotypen (wie ein Zugangskontrollsystem, das annimmt, dass alle Menschen mit Doktor-Titel männlich sind¹¹).

Auch neue Polizei-Technologien könnten einen solchen Effekt haben: Die Polizei in Chicago setzt maschinelles Lernen ein, um aus Listen registrierter StraftäterInnen neue Listen von bisher nicht auffällig gewordenen Menschen zu generieren, die ein (angeblich) erhöhtes Risiko haben, straffällig zu werden. Dazu werden Faktoren wie die Kriminalitätsrate der Nachbarschaft, aber auch die sozialen Netzwerke der Personen verwendet. Die derart identifizierten Menschen werden von der Polizei darüber informiert, dass man ein Auge auf sie hat.¹² Doch was passiert, wenn bestimmte Minderheiten wie Afro-Amerikaner in diesen Strafregistern überrepräsentiert sind, weil die Polizei durch *racial profiling* überdurchschnittlich viele von ihnen wegen Drogendelikten festnimmt? Das scheint wie ein Einfallstor für *preexisting bias* in diese angeblich neutralen Algorithmen.

Bisher laufen solche Programme noch nicht lange genug, um diese Befürchtungen zu bestätigen oder zerstreuen. Doch die Erfahrung zeigt: Wenn eine Technologie erst einmal im Einsatz ist, ist es ungleich schwerer, sie fundamental zu ändern oder gar wieder abzuschaffen, falls sich herausstellt, dass sie schädlich ist.

Wäre es nicht besser, solche Fragen zu klären, bevor ein System in den Produktiveinsatz geht?

Bei vielen Algorithmen liegt die Fehler- und Verfälschungsfreiheit sogar im Interesse der Hersteller – wenige Firmen bauen absichtlich rassistische Algorithmen, und noch weniger möchten öffentlich des Rassismus' bezichtigt werden. Insofern fehlen meist nicht die Gesetze, sondern die Methoden, um festzustellen, ob ein Algorithmus verzerrend wirkt oder nicht – und das Bewusstsein, dass dieses Problem überhaupt existiert.

Das Thema der Fairness von Algorithmen findet sich in kaum einem Lehrplan. Viele Informatiker:innen gehen ins Berufsleben, ohne auch nur von diesem Problem gehört zu haben. Dadurch fehlt ihnen der kritische Blick, um ihre Algorithmen und Systeme auf diese Probleme zu prüfen. Und während sexistische Zugangskontrollsysteme leicht erkennbar sind, ist in Zeiten von Big Data ein *preexisting bias* in der Datenbasis nicht leicht aufzudecken.

Perspektive

Schon heute entscheiden Algorithmen über die Kredite, die wir erhalten (oder auch nicht), die Suchergebnisse, die wir sehen (oder auch nicht), und die Status-Updates unserer Freunde, die uns unser *social network* anzeigt (oder auch nicht). Deswegen ist es wichtig, die Frage nach der Fairness dieser Algorithmen zu stellen und zu beantworten.

Die Instrumente existieren: Friedman *et al.* vertreten den Ansatz des *value-sensitive design*,¹³ bei dem schon während des Design-Prozesses über die möglichen sozialen Folgen und Nachteile der Design-Entscheidungen nachgedacht werden soll. Shneiderman *et al.* formalisieren dieses Verfahren mit so genannten *social impact statements*, die Entwickler dazu bringen, strukturiert schon während der Entwicklung über mögliche negative Auswirkungen von Algorithmen nachzudenken und diese zu beheben.¹⁴ Auch Ansätze wie *Reflective Design*¹⁵ von Phoebe Sengers *et al.* enthalten Ideen, die für das Erkennen und Entfernen von Vorurteilen und Verfälschungen in Algorithmen hilfreich sein können. Wir sollten uns mit dieser Problematik vertraut machen, wenn wir möchten, dass unsere Arbeit ein Teil der Lösung anstatt ein Teil des Problems wird.

Lawrence Lessig schreibt „*Code is Law*“.¹⁶ Angesichts der zunehmenden Macht von Algorithmen liegt es auch an uns Informatiker:innen, sicherzustellen, dass diese Gesetze nicht diskriminierend sondern *fair* sind.

Max Maaß



Max Maaß studiert IT-Sicherheit in Darmstadt und forscht im Bereich der *Privacy-Enhancing Technologies*.

Anmerkungen

- 1 Predictive Policing wird hier als Oberbegriff für eine Reihe verschiedener Verfahren genutzt, von der automatischen Planung von Polizeistreifen basierend auf Kriminalstatistiken bis hin zur Risikobewertung, welche BürgerInnen eventuell straffällig werden könnten.
- 2 Im Englischen wird der Begriff *bias* verwendet, der alles von unabsichtlicher Verfälschung bis zu bewusstem Rassismus bezeichnen kann. Im Deutschen gibt es kein äquivalentes Wort.
- 3 L. G. Locke, "Flying the Unfriendly Skies: The Legal Fallout Over the Use of Computerized Reservation Systems as a Competitive Weapon in the Airline Industry", *Harv. J. Law Technol.* Vol. 1449, pp. 219-237, 1990
- 4 S. Lowery, "A Blot on the Profession", *Br. Med. J.*, vol 296, no. 6623, pp. 657-658, 1988.
- 5 C. Huff and J. Cooper, "Sex Bias in Educational Software: The Effects of Designers' Stereotypes on the Software They Design", *J. Appl. Soc. Psychol.*, vol. 17, no. 6, pp. 519-532, 1987.
- 6 B. Friedman, E. Brok, S. K. Roth, and J. Thomas, "Minimizing Bias in Computer Systems", *ACM SIGCHI Bull.*, vol. 28, no. 1, pp. 48-51, 1996.
- 7 P. Sengers, K. Boehner, S. David, and J. "Joyfish" Kaye, "Reflective Design", *Proc. 4th Decenn. Conf. Crit. Comput. Between Sensib.*, pp. 49-58, 2005.
- 8 Siehe zum Beispiel M. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings", *Proc. Priv. Enhancing Technol.*, vol. 2015, no. 1, pp. 92-112, 2015 and L. Sweeney, "Discrimination in Online Ad Delivery", *Commun. ACM*, vol. 56, pp. 44-54, 2013.
- 9 N. Diakopoulos, "Algorithmic Accountability Reporting: On the Investigation of Black Boxes", *Tow Cent. Digit. Journal.*, 2014.
- 10 T. Z. Zarsky, "Understanding Discrimination in the Scored Society", *Washingt. Law Rev.*, vol. 89, no. 4, pp. 1375-1412, 2014.
- 11 <http://www.dailymail.co.uk/news/article-3000705/Female-paediatrician-locked-gym-changing-room-computer-assumed-Doctor-title-MAN.html>
- 12 <https://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist>
- 13 B. Friedman, P. H. Kahn Jr., and A. Borning, "Value Sensitive Design and Information Systems," *Human-Computer Interact. Manag. Inf. Syst. Found.*, pp. 348-372, 2006.
- 14 B. Shneiderman and A. Rose, "Social impact statements," *Proceedings of the Symposium on Computers and the Quality of Life*, pp. 90-96, 1996.
- 15 P. Sengers, K. Boehner, S. David, and J. "Joyfish" Kaye, "Reflective Design", *Proc. 4th Decenn. Conf. Crit. Comput. Between Sensib.*, pp. 49-58, 2005.
- 16 <http://harvardmagazine.com/2000/01/code-is-law-html>



Dagmar Boedicker

Refugees Welcome – at Your Secret Service

Zur Zeit helfen viele Ehrenamtliche den Flüchtlingen. Ohne ihr freiwilliges Engagement würde vielerorts kaum etwas laufen. Sie betreuen unbegleitete Minderjährige oder Familien, geben Deutschunterricht, machen Stadtführungen, übernehmen Patenschaften, ... Nur haben sie leider kaum eine oder gar keine Sensibilität für die Notwendigkeit, die Daten der Flüchtlinge zu schützen. Datenschutz für Flüchtlinge? – Fehlanzeige!

Deutschland hat sich viele Jahre lang weggeduckt. Wir liegen in der Mitte der EU, umgeben von anderen Mitgliedstaaten, die sich nach dem Dublin-Abkommen um Aufnahme von und Asyl für geflohene Menschen kümmern mussten. Seit das UNHCR eigentlich gar kein Geld mehr für den Anteil der 60 Millionen¹ gewaltsam Vertriebener hat, um die es sich kümmern soll und will, findet eine kleine Völkerwanderung in die EU statt. Dafür sind wir und andere Industriestaaten verantwortlich, auch, aber nicht nur, weil wir versprochene Zahlungen an die UN nicht geleistet haben.²



Migranten am Wiener Westbahnhof am 5. September 2015, die zu Tausenden Richtung Deutschland weiterreisen

Foto: Bwag, CC BY-SA 4.0

Jetzt ist es kompliziert, eine menschenwürdige Unterbringung und Betreuung so vieler Menschen möglichst schnell zu organisieren. Ehrlicherweise müssen wir aber auch zugeben, dass es keine Regierung geschafft hätte, ausreichend Unterkünfte, medizinische und psychologische Hilfe, Bildungsangebote usw. auf Vorrat zu halten. Der Protest gegen eine solche *Mittelverschwendung* wäre Pegida-ähnlich gewesen. Dasselbe gilt für den Versuch, Fluchtursachen zu bekämpfen: Prävention ist nicht vermittelbar, nur mit Krisenbewältigung kann Politik Anerkennung ernten.

Information und Kommunikation

Für ein Land mit 80 Mio. Einwohnern ist eine Million Flüchtlinge sicher keine Katastrophe. In München aber ließ sich beobachten, dass Information und Kommunikation große Schwierigkeiten bereiten. Selbst die Profis in der Verwaltung oder den caritativen Einrichtungen sind überlastet und schaffen es nicht immer, Koordination und Informationsvermittlung aktuell zu halten.

Wie lassen sich die Ehrenamtlichen vernetzen? Wie können sie ihre Termine abstimmen? Natürlich über Twitter, Google oder Facebook! Anfang September führte ein Tweet von Helfer:innen am Hauptbahnhof (*Wir haben keine Milch mehr!*) dazu, dass