



Brauchen wir eine Rechenschaftspflicht für algorithmische Entscheidungen?

Kurzfassung eines Artikels von Nicholas Diakopoulos

Nicholas Diakopoulos · Oliver Deussen

Jedes Quartal erzeugen automatisierte Schreibalgorithmen Tausende von Artikeln über Unternehmensgewinne für Associated Press (AP), die auf nicht viel mehr als nur strukturierten Daten aufbauen. Unternehmen wie Automated Insights, das diese Artikel für AP herstellt, können mittlerweile Nachrichtenbeiträge für fast jeden Bereich schreiben, der saubere und gut strukturierte Daten aufweist: darunter natürlich das Finanzwesen, aber auch Gebiete wie Sport, Wetter und Bildung. Diese Beiträge sind keineswegs mühsam oder schwer lesbar; sie sind anschaulich geschrieben, legen einen individuell klingenden Ton und Stil an den Tag und machen es manchen Lesern sogar schwer, sie von Beiträgen zu unterscheiden, die von Menschenhand geschrieben wurden [4].

Die algorithmische Kuratierung von Inhalten, also das automatisierte Zusammentragen und Sichten von Inhalten steckt auch hinter einigen der wichtigsten und einflussreichsten Nachrichtenplattformen, die wir fast alle nutzen. Nach einer aktuellen Studie benutzt Facebook rund 61 % seiner Nutzer der Generation Y als Nachrichtenquelle zu Themen wie Regierung und Politik [13], wobei der Öffentlichkeit nicht bewusst ist, dass der Facebook-Newsfeed von Algorithmen kuratiert wird [9]. Das ist besonders deshalb problematisch, weil Facebook das Wahlverhalten seiner Nutzer allein durch die Anzahl der „hard news“ (Informationen über aktuelle Ereignisse aus Politik, Gesellschaft, Wirtschaft und Natur) beeinflussen kann, die in einem individuellen Newsfeed promotet werden [16]. Diese Information wird von neuesten Forschungsergebnissen untermauert, die nahelegen, dass verzerrte Suchergebnisse die Vorlieben

von noch unentschiedenen Wählern beeinflussen können [8]. Dies verdeutlicht, wie wichtig es ist, zu hinterfragen, inwieweit solche Kuratierungen und Rangordnungen demokratische Prozesse beeinflussen.

Dies sind nur ein paar Beispiele dafür, wie Algorithmen unsere Wahrnehmung von Medien und Informationen beeinflussen. Es ist an der Zeit, ernsthaft darüber nachzudenken, wie algorithmisch gesteuerte Entscheidungen, die große Teile der Gesellschaft betreffen, öffentlich verantwortlich für ihr Handeln gemacht werden können. Gerade im Hinblick auf teure Fehler, auf Diskriminierung, Zensur oder die unfaire Verweigerung öffentlicher Zuwendungen stellt sich die Frage, wann und wie Algorithmen zu regulieren sind. Informatiker und Ingenieure spielen hier eine wichtige Rolle. Autonome Entscheidungsprozesse mögen zwar das Kernstück algorithmischer Macht bilden, es gibt jedoch jede Menge menschlicher Einflüsse auf diese Algorithmen: die Auswahl von Kriterien, die Funktionen zu ihrer Optimierung, die Art der Trainingsdaten oder der genutzten Kategorien, um hier nur einige zu nennen.

Algorithmische Entscheidungen

Zunächst sollten die verschiedenen Arten von Entscheidungen nachvollzogen werden, die Al-

DOI 10.1007/s00287-017-1059-9
© Springer-Verlag Berlin Heidelberg 2017

Nicholas Diakopoulos ist Assistenzprofessor an der University of Maryland. Er ist außerdem Mitglied des Tow Center for Digital Journalism der Columbia University.

Übersetzung und Zusammenfassung:
Prof. Dr. Oliver Deussen (Universität Konstanz)

Kuratieren vs. Recherchieren

Der Begriff Kuratieren hat seine angestammte Bedeutung zunächst aus der Kunst. Man kuratiert Ausstellungen, d. h. man konzipiert zu einem Thema eine Ausstellung und legt die erforderlichen Kunstwerke fest. Danach wird versucht diese einzusammeln und auszustellen. Davon abweichend hat in den letzten Jahren der Begriff im Internet eine neue Bedeutung erhalten, die v. a. im Online-Journalismus eine große und wachsende Rolle spielt: das algorithmengesteuerte Sammeln und Gewichten von Informationen im Internet, die nicht mehr recherchiert werden. Bedeutung erlangt die kuratierte Information etwa durch die Anzahl von Ergebnissen, die dann von Suchmaschinen dazu ausgegeben werden oder durch die Anzahl an „likes“ oder Followern, die ein kuratierter Trend anzieht.

gorithmen treffen. Dazu gehören Priorisierungs-, Klassifizierungs-, Zuordnungs- und Filterprozesse:

Wir *priorisieren* jeden Tag, um mit der Informationsflut umzugehen. Algorithmen machen dasselbe, sie priorisieren Informationen dahingehend, dass sie einige Informationen zulasten von anderen in den Vordergrund stellen. Priorisierung hat aber auch immer mit Selektion und Diskriminierung zu tun. Suchmaschinen sind gute Beispiele dafür, aber auch viele andere wichtige Rankings, etwa zur Qualität von Hochschulen oder Krankenhäusern, oder solche, die das Gefährdungspotenzial von Einwanderern in Beobachtungslisten darstellen [11]. Die Kriterien, die in solchen Rankings zum Einsatz kommen, ihre Definition und Repräsentation in Daten, ihre Gewichtung stellen grundlegende Designentscheidungen dar, die immer wieder gründlich überdacht und geprüft werden müssen.

Klassifizierungen weisen etwas oder jemanden, also z. B. einen Menschen, einer bestehenden Klasse zu, indem sie besondere Merkmale hervorheben. Die Zugehörigkeit zu einer Klasse steuert danach eine Vielzahl weiterer Entscheidungen. Die Trainingsdaten für die lernenden Algorithmen spielen hierbei eine wichtige Rolle, insbesondere im Hinblick auf menschliche Präferenzen, die darin verborgen sein können. Neue Forschungsergebnisse zeigen, wie

wichtig es ist, den Ausschnitt der Gesellschaft genau zu betrachten, aus dem Trainingsdaten erhoben werden [15].

Bei der *Zuordnung* werden Beziehungen zwischen Einheiten hergestellt. Hierbei werden sehr unterschiedliche Beziehungen verwendet, ganz grundsätzliche wie verwandt mit oder ähnlich zu oder besondere mit domänenspezifischen Bedeutungen. Diese Zuordnungen führen dann zu ganz unterschiedlichen Annahmen bei der Auslegung durch den Menschen. So gab es den Fall eines Mannes, der in Deutschland nach seinem Namen googelte und durch die Autovervollständigungsfunktion mit den Schlagworten Scientology und Betrug in Verbindung gebracht wurde. Der Betroffene strengte daraufhin einen Verleumdungsprozess an, den er letztlich auch gewann [5].

Ein großes Problem beim Arbeiten mit großen Datenmengen ist unser Glaube an die Korrelation. Eine Korrelation ist erst einmal nicht mehr als eine statistische Beziehung zwischen Datendimensionen. Aber obwohl wir es besser wissen müssten, werden Korrelationen allzu oft als kausale Ursachen verstanden. So mag der Mann, der auf Google mit Betrug in Verbindung gebracht wurde, Grund für diese Zuordnung geliefert haben oder eben auch nicht. Aber wir legen es so aus.

Beim *Filtern* geht es darum, Informationen aufgrund verschiedenster Kriterien zu berücksichtigen oder auszuschließen. Dies kommt oft an der Schnittstelle mit dem Nutzer zum Tragen, etwa bei Nachrichten-Apps wie Facebook oder Flipboard. Wegen vieler vorhandener Trolle oder destruktiver Nutzer spielen Filter beim Publizieren in sozialen Medien eine große Rolle. Online verfasste Kommentare werden algorithmisch gefiltert, um festzustellen, ob sie gesellschaftsfeindliche Inhalte transportieren und deswegen nicht der Öffentlichkeit zugänglich gemacht werden sollten. Natürlich besteht hierbei die Gefahr, zu weit zu gehen und in Zensur abzugleiten.

Im Hinblick auf die vielen Entscheidungen und Vorhersagen, die Algorithmen treffen, insbesondere jene, die Menschen oder Eigentum betreffen, ist es daher äußerst wichtig, ethische Regeln der Softwaretechnik zu beachten. Der Code of Ethics für Softwareingenieure der Association for Computing Machinery (ACM) listet acht Prinzipien des professionellen Verhaltens auf [2]: Zuallererst sollen Softwareingenieure im öffentlichen Interesse han-

deln, sie sollen für ihre Arbeit verantwortlich sein und zur Rechenschaft gezogen werden können, sie sollen private Interessen gegen öffentliches Wohl abwägen, Sicherheit und Privatsphäre gewährleisten, Täuschung vermeiden und an Benachteiligte denken. Die generellen moralischen Regeln, die von ACM aufgestellt wurden, beinhalten auch das Gebot, „Schaden gegenüber anderen zu vermeiden“, „fair zu sein und darauf zu achten, nicht zu diskriminieren“ und „das Recht des Anderen auf Privatsphäre zu achten“.

Aber eigentlich wird noch viel mehr benötigt: Ethische Ideale müssen den gesamten Programmierungsprozess bestimmen. Softwareingenieure müssen sich ständig fragen: Was sind die Konsequenzen eines unwahrscheinlichen, aber vielleicht fatalen falsch präsentierten Ergebnisses oder die Auswirkungen davon, wie die Kriterien in den verwendeten Trainingsdaten definiert werden?

Rechenschaftspflicht in der Regierung und im privaten Sektor

In einem modernen demokratischen Staat wählen die Bürger eine Regierung, die soziale Leistungen erbringt und deren Machtausübung und Kontrollfunktionen durch Normen und Gesetze reguliert werden. Die Regierung ist nur insoweit legitim, als dass sie gegenüber den Bürgern zur Rechenschaft gezogen werden kann. Algorithmen sind fast vollständig unreguliert und sie üben tatsächlich eine Art von Macht über Individuen oder politische Prozesse aus, die in einigen Fällen (z. B. geheime Beobachtungslisten der Regierung) überhaupt keinen Rechenschaftszwängen unterliegt. Eine jüngst durchgeführte Überprüfung von Verwaltungsmodellen in der Social Security Administration des amerikanischen Sozialversicherungssystems, die benutzt werden, um Lebenserwartung und Zahlungsfähigkeit vorherzusagen, zeigt, dass beide Parameter systematisch unterschätzt wurden, was zu der Annahme führte, dass mehr Mittel zur Verfügung standen, als tatsächlich der Fall war [12]. Angesichts der wichtigen politischen Entscheidungen, die von solchen Vorhersagen beeinflusst werden, sollten wir, die Regierten, die fehlende Toleranz nicht tolerieren, auch nicht die Durchführung systematischen Benchmarkings oder die Auswertung von Vorhersagen. Deswegen sollten wir Transparenz bei unseren Regierungen einfördern. Eine totale Transparenz bis hin zum Quellcode

ist dabei in den meisten Fällen gar nicht nötig. Oft wäre es viel sinnvoller, nur die Herausgabe wichtiger Informationen zu fordern, etwa von gesammelten Ergebnissen und verwendeten Benchmarks.

Bei Unternehmen gelten andere Maßstäbe, obwohl sie manchmal durch sozialen Druck zum Handeln gezwungen werden (z. B. durch Boykotte). Noch schlagkräftiger mag hier das Argument sein, dass eine bessere Datenqualität und damit bessere Rückschlüsse zu größerer Kundenzufriedenheit führen. Um das zu erreichen, ist es am besten, Prozesse so zu gestalten, dass dem Endanwender die Korrektur von falschen Ergebnissen ermöglicht wird. Wenn beispielsweise Kreditratingagenturen ihre Kriterien zur Bewertung von Individuen veröffentlichen würden, wäre es dann nicht sehr hilfreich, wenn jeder seine Bewertung beeinflussen könnte? In diesem Spiel würden die Agenturen dann auch gezwungen, verantwortungsvoller zu handeln.

Ein Transparenzstandard für Algorithmen

Was genau kann und soll nun über Algorithmen publik gemacht werden? Auf der Grundlage eines von mir veranstalteten Workshops zu diesem Thema wurden fünf breit gefasste Informationskategorien aufgestellt, die uns zur Veröffentlichung geeignet erscheinen: menschliche Beteiligung, Daten, das Modell, Rückschlüsse und Einsatz.

Menschliche Beteiligung. Ziel, Absicht, und Zweck eines Algorithmus müssen erklärt werden, inklusive der redaktionellen Ziele. Wer in einem Unternehmen hat direkte Kontrolle über den Algorithmus? Wer ist verantwortlich? Wie in jedem kollektiven Arbeitsprozess wird es hier schwierig sein, punktgenau festzuhalten, wer genau was beigetragen hat (oder wer genau für welchen Fehler verantwortlich ist; [14]). Doch die Offenlegung menschlicher Beteiligung würde nichtsdestotrotz soziale Effekte erzeugen, die sich einerseits am Ruf der beteiligten Individuen festmachen und es andererseits Trittbrettfahrern schwerer machen. Beteiligte würden gegebenenfalls mehr Verantwortung gegenüber der Öffentlichkeit verspüren, wenn ihr Name mit einem Algorithmus verbunden ist.

Daten. Es gibt viele Möglichkeiten, Transparenz über die Daten zu erzeugen, die Algorithmen zugrunde liegen. Die Qualität der Daten könnte kommuniziert werden, z. B. ihre Genauigkeit, ihre

Vollständigkeit, ihre Zeitlichkeit (schließlich kann sich die Geltung von Daten mit der Zeit ändern), die Repräsentativität von Trainingsdaten für eine bestimmte Bevölkerungsgruppe sowie andere Annahmen und Einschränkungen. Wie wurden Daten definiert, gesammelt, umgewandelt, verifiziert und bearbeitet? Welche Arten persönlicher Informationen werden genutzt, wie sehen gesammelte oder per Rückschluss erstellte Profile von Individuen aus; wer hat Personalisierungen vorgenommen?

Modell. Das Modell sowie der Modellierungsprozess könnten ebenfalls bis zu einem gewissen Grad transparent gestaltet werden. Auf welche Merkmale oder Variablen stützt sich ein Algorithmus? Oft werden solche Merkmale gewichtet: Wie genau sehen die Gewichtungen aus? Welche Werkzeuge wurden im Modellierungsprozess genutzt? Aufgrund welcher menschlicher Annahmen wurden Gewichtungen vorgenommen, wie hat sich der Designprozess gestaltet, wurden alternative Modelle in Betracht gezogen, wurde zwischen diesen Modellen verglichen? Was sind die statistischen oder sonstigen Annahmen, auf denen ein Modell fußt, und woher kommen diese Annahmen? Und falls ein Aspekt des Modells im Frontend nicht sichtbar gemacht wurde: warum war das so?

Rückschlüsse. Die Rückschlüsse, die ein Algorithmus zieht, etwa Klassifizierungen oder Prognosen, geben oft Anlass, an seiner Genauigkeit zu zweifeln und sie bergen Spielräume für Fehler. Die Hersteller von Algorithmen könnten daher den Abgleich mit Standarddatensätzen und -maßnahmen zur Bewertung der Genauigkeit in Erwägung ziehen und Schlüsselstatistiken veröffentlichen. Wie hoch ist der Spielraum für Fehler? Welche Schritte werden unternommen, um bekannte Fehler zu beheben? Sind diese Fehler auf menschliche Beteiligung zurückzuführen, auf Dateninput oder auf den Algorithmus selbst? Klassifikatoren erzeugen oft einen Vertrauenswert, dieser könnte veröffentlicht werden als Maßstab für die Unsicherheiten im Ergebnis.

Einsatz. Wann wurde ein Algorithmus überhaupt eingesetzt? Welche Elemente einer Ergebnismenge wurden weggefiltert? Im Fall von Facebook etwa sollte man sich fragen, was man nicht sieht, oder andersherum, welche eigenen Posts (z. B. in einem Newsfeed) von anderen nicht gesehen werden. Tech-

nische Systeme sind ständig in Bewegung, deshalb muss jeder Versuch der Offenlegung immer die Dynamik in Betracht ziehen, der Algorithmen unterliegen, die Tag für Tag von neuen Daten lernen. Die Ingenieurskultur muss sich die Idee einer ständigen Beurteilung zu eigen machen. Vielleicht müssten auch neue, interdisziplinäre Berufe geschaffen werden, die sich ausschließlich mit dem Testen von algorithmischen Risiken oder von algorithmischer Transparenz befassen.

Wir sollten außerdem Vorschriften in Betracht ziehen, die die Herausgabe von Informationen im Zusammenhang mit algorithmisch gesteuerten Entscheidungen vorsehen, wie etwa bei Kreditwürdigkeitsprüfungen [3]. In anderen Fällen ist u. U. auch ein konfliktbetontes Vorgehen nötig, etwa bei der Untersuchung von Blackbox-Algorithmen. Im journalistischen Kontext nenne ich das „algorithmic accountability reporting“ [6]. Hier wird anhand von Stichproben algorithmischer Schlüsseldimensionen das Verhältnis von Input zu Output untersucht, um die Fehler oder Voreingenommenheit eines Algorithmus zu ermitteln und zu beschreiben.

Ein Beispiel: Im Frühjahr 2015 untersuchte ich den „surge-pricing algorithm“ (Preissprunzalgorithmus) bei Uber [7]. Die App benutzt dynamische Preise, um mehr Fahrer dazu zu bewegen, online zu gehen und versucht auf diese Weise, Angebot und Nachfrage aneinander anzugleichen. Das ist leicht nachvollziehbar und steht im Einklang mit grundlegenden ökonomischen Theorien. Meine Studie, die auf Daten zur Preisgestaltung fußte, die innerhalb eines Monats in Washington, D.C. erhoben worden waren, legt aber nahe, dass durch die Preissprungpolitik nicht mehr Fahrer dazu bewegt wurden, aktiv zu werden, sondern bereits aktive Fahrer einfach umverteilt. Das ist wichtig, weil es bedeutet, dass sich die vorhandenen Fahrer eher in solche Viertel begaben, in denen höhere Preise zu erzielen waren. Das führte dazu, dass andere Viertel vernachlässigt wurden. Die Fahrzeuge von Uber stellen rivalisierende Güter dar. Meine Analyse regt vielleicht dazu an, darüber nachzudenken, in welchen Vierteln es zu besserer oder schlechterer Servicequalität kommt. Höhere Preise und besserer Service an einer Stelle heißt schlechterer Service für andere.

Fazit

Die Gesellschaft muss sich dringend mit dem Einfluss beschäftigen, den Algorithmen in Wirtschaft

und öffentlicher Verwaltung haben, sodass entsprechende Mechanismen zur Offenlegung der verwendeten Parameter von Anfang an in diese Systeme integriert werden. Mir scheint es wichtig, dass die hier beschriebenen Ideen zum ethischen und verantwortungsvollen Umgang mit Algorithmen in die Praxis übernommen werden, insbesondere wenn diese mit der Fähigkeit ausgestattet werden, Entscheidungen zu treffen. Es muss noch viel Forschung betrieben werden, um die entsprechenden Dimensionen und Modalitäten algorithmischer Transparenz zu verstehen, um zu lernen, wie interaktive Modelle eingebunden werden können, wie der Journalismus sich weiterentwickeln muss und wie man Softwareingenieure für diese Probleme sensibilisieren und entsprechend schulen kann.

Originalartikel

1. Diakopoulos N (2016) Accountability in Algorithmic Decision Making. *Communications of the ACM* 59(2):56–62. DOI 10.1145/2844110

Literatur

2. ACM (2015) Software Engineering Code of Ethics and Professional Practice, <https://www.acm.org/about/se-code#full>

3. Citron DK, Pasquale F (2014) The scored society: due process for automated predictions. *Wash Law Rev* 89(1):1–33
4. Clerwall C (2014) Enter the robot journalist. *Journalism Practice* 8(5):519–531
5. Diakopoulos N (2014) Algorithmic defamation: the case of the shameless auto-complete. *Tow Center for Digital Journalism*
6. Diakopoulos N (2015) Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3(3):398–415
7. Diakopoulos N (2015) How Uber surge pricing really works. *Washington Post Wonkblog* (April 17, 2015)
8. Epstein R, Robertson RE (2015) The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *PNAS* 112(33):E4512–E4521
9. Eslami M et al. (2015) „I always assumed that I wasn't really that close to [her]“: Reasoning about invisible algorithms in the news feed. In: *Proceedings of the 33 Annual ACM SIGCHI Conference on Human Factors in Computing Systems*
10. Herlocker JL et al (2000) Explaining collaborative filtering recommendations. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp 241–250
11. Kalhan A (2013) Immigration policing and federalism through the lens of technology, surveillance, and privacy. *Ohio State Law J* 74(6):1105–1165
12. Kashin K et al. (2015) Systematic bias and nontransparency in US Social Security Administration forecasts. *J Econ Perspect* 29(2):239–258
13. Mitchell A et al. (2015) Millennials and Political News. *Pew Research Center, Journalism and Media* (June 1, 2015), <http://www.journalism.org/2015/06/01/millennials-political-news/>
14. Nissenbaum H (1996) Accountability in a computerized society. *Sci Eng Ethics* 2(1):25–42
15. Sen S et al. (2015) Turkers, Scholars ‚Arafat‘ and ‚Peace‘: Cultural communities and algorithmic gold standards. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp 826–838
16. Sifry M (2014) Facebook wants you to vote on Tuesday. Here's how it messed with your feed in 2012. *Mother Jones* (Oct. 31, 2014), <http://www.motherjones.com/politics/2014/10/can-voting-facebook-buttonimprove-voter-turnout>