Noémie Ammann / Remo R. Schmidlin*

Towards an Algocracy?

Tagungsbericht zur interdisziplinären Tagung «Towards an Algocracy? – Interdisciplinary Approaches to Algorithm Governance», vom 9. November 2018, organisiert von Prof. Dr. Isabelle Wildhaber und Prof. Dr. Melinda Lohmann.

I. Einleitung

Die Digitalisierung unserer Gesellschaft, unseres Arbeitsplatzes und unserer Freizeit hat ein Niveau erreicht, das nicht mehr zu ignorieren ist. In seiner Eröffnungsansprache unterstreicht Prof. Dr. Markus Müller-Chen als Dekan der Law School an der Universität St. Gallen die Problematik im Umgang mit Algorithmen durch den Tagungsslogan: Algocracy – Algocrazy – I'll go crazy! Es findet sich kein Konsens darüber, wie mit Algorithmen und künstlicher Intelligenz (KI) umgegangen werden soll, wenn diese unsere menschliche Autonomie tangieren. Die Gegensätzlichkeit zwischen noch nie dagewesenem Komfort einerseits und der Herausforderung einer sinnvollen und wirksamen *Algorithm Governance* andererseits steht stets im Kern der Debatte.

Deshalb wirft Prof. Dr. Isabelle Wildhaber eingangs die Frage auf, ob wir auf dem Weg zu einer *Algocracy* sind, d.h. einer Welt, in der unsere Gesellschaft von Algorithmen bestimmt ist und in der der Mensch es nicht (mehr) vermag, diese zu regulieren.

Die Diskriminierung durch Algorithmen ist dabei eine der gegenwärtig grössten Herausforderungen. Wie von Prof. Dr. Melinda Lohmann präsentiert, gibt es zahlreiche Beispiele, die das illustrieren. Algorithmen sind nicht per se objektiv, sondern werden von menschlichen Vorurteilen geprägt: «Data and data sets are not objective; they are creations of human design.»[1] Umso wichtiger ist es, die *Black Box* von Algorithmen so weit wie möglich zu enthüllen und zu verstehen.

Trotz der Komplexität der Thematik und Undurchschaubarkeit von Algorithmen und ihres Einflusses auf unsere Gesellschaft ist es unter den Referierenden unbestritten, dass eine interdisziplinäre Diskussion essenziell ist, um eine allfällige *Algorithm Governance* auszuarbeiten.

II. Soziologie einer algorithmisierten Welt

Der Prozess der Digitalisierung lässt sich am besten durch eine abstrakte Betrachtungsweise verstehen. Algorithmen und digitale Kulturen sind im Endeffekt nichts anderes als Software Engineering, schildert PD Dr. Jan-Hendrik Passoth (Technische Universität München). Dabei sind Software, Daten und Algorithmen die Bausteine von digitalen Kulturen.

Ihm zufolge haben sich in den letzten Jahrzehnten zwei Ansätze herauskristallisiert, welche darauf abzielen, die Einflussnahme von Algorithmen auf moderne Gesellschaften zu verstehen. Einerseits soll in der Öffentlichkeit ein Verständnis für die Funktionsweise von Algorithmen geschaffen werden (sog. *Code Literacy*). Andererseits wird versucht, durch das Öffnen der *Black Box* von Algorithmen die einem System inhärente menschliche Komponente – der sogenannte *Mechanical Turk* – zu extrahieren und so Einblicke in die Funktionsweise von Algorithmen zu gewinnen. Gemäss Passoth stehen moderne Gesellschaften vor der Schwierigkeit, ein Konzept zu entwickeln, das es vermag, zwischen den beschriebenen Ansätzen zu vermitteln. Mit anderen Worten soll ein Gleichgewicht zwischen der Hermeneutik von Software und dem Reiz, das *Human Bias* zu entdecken, hergestellt werden. In dem Sinne appelliert Passoth in den Worten von Ian Bogost an eine pragmatischere Herangehensweise: «Let's bring algorithms down to earth again. Let's keep the computer around without fetishizing it.»[2]

Abschliessend analysiert Passoth digitale Gesellschaften, welche als Kulturen der Berechnung angesehen werden können. Der Prozess der Berechnung besteht darin, Informationen zu dekontextualisieren, sie auf Grundlage von Regeln und Prinzipien zu transformieren, um sie anschliessend wieder zu rekontextualisieren, d.h. in das Ergebnis einzubetten. Die Transformation von Daten erfordert jedoch in einem gewissen Masse einen institutionellen Rahmen. An dieser Stelle ist die Gleichung um den Faktor Politik zu ergänzen, was Interventionen und Alternativen ermöglicht, indem verschiedene Interessengruppen in den Review-Prozess einbezogen werden.

III. Algorithmische Transparenz und Datenschutz

Prof. Lilian Edwards (University of Strathclyde, Glasgow) beginnt ihren Beitrag mit einem Überblick über jüngste Ereignisse, die für Aufsehen in der Öffentlichkeit gesorgt haben und im Zusammenhang mit automatisierten Entscheidungssystemen stehen. Im Jahr 2016 enthüllte ProPublica, dass das in den USA häufig angewandte Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) verzerrte Ent-

* Noémie Ammann, B.A. HSG, und Remo R. Schmidlin, B.A. HSG, beide Assistierende am Forschungsinstitut für Arbeit und Arbeitswelten der Universität St. Gallen. Alle Internetquellen besucht im September 2019.

[1] Kate Crawford, The Hidden Biases in Big Data, Harvard Business Review 1. April 2013, <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

[2] Ian Bogost, The Cathedral of Computation, The Atlantic 15. Januar 2015, <https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/>.

scheidungen zum Nachteil von Bürgern dunkler Hautfarbe trifft.[3] Darüber hinaus wird Algorithmen vorgeworfen, demokratische Grundprinzipien ernsthaft zu gefährden, was sich insbesondere anhand des Cambridge-Analytica-Skandals zeigt. Dieser medialen Aufmerksamkeit ist zu entnehmen, dass besonders *Machine Learning*-Algorithmen unsere Gesellschaft in mancher Hinsicht vor komplexe Fragen stellen. Im Kern der Funktionsweise von heute gebräuchlichen Algorithmen steht die Replikation von Daten. Werden grosse Datenmengen zum Training in ein System eingespeist, übernimmt dieses unweigerlich auch die bestehenden Verzerrungen aus den Trainingsdaten.

Edwards beschäftigt sich mit der Frage, wie das Datenschutzrecht mit diesem Problemfeld umgeht. In der Europäischen Union ist das «Recht auf Erklärung» ein oft genanntes Rechtsmittel im Zusammenhang mit der DSGVO.[4] Mögliche Anspruchsgrundlage ist dabei Art. 22 DSGVO, der jedoch aufgrund seiner restriktiven Formulierung stark eingeschränkt ist. Entgegen der viel verbreiteten Ansicht nennt Art. 22 kein «Recht auf Erklärung». Edwards erklärt sich dieses Missverständnis mit der kurzen Erwähnung eines «Rechts auf Erklärung» in den rechtlich nicht bindenden Erwägungsgründen.[5] Generell scheinen die Erwägungsgründe in vielen Fällen im Widerspruch zum Haupttext zu stehen. Dies veranschaulicht, dass sich die DSGVO als individuelle Anspruchsgrundlage nur beschränkt eignet und hauptsächlich durch den ausgiebigen und immer noch anhaltenden politischen Diskurs geprägt ist. In Anbetracht der Technophobie und des Mangels an Macht und Ressourcen der betroffenen Personen fordert Edwards mehr systemische und nicht-verbraucherbasierte Rechtsmittel. Dies erfordert, dass der Gesetzgeber andere Rechtsinstrumente wie das Arbeits- oder Gleichstellungsrecht, die gerichtliche Überprüfung oder die Informationsfreiheit in Betracht zieht. Tatsächlich stellt die DSGVO bereits heute nicht-individuelle Ansätze bereit, wie beispielsweise Datenschutz-Folgenabschätzungen[6] oder Anforderungen an die Ausgestaltung von automatisierten Entscheidungssystemen, wie *Privacy by Design*[7] und *Privacy by Default*[8]. Edwards fordert dazu auf, die Weiterentwicklung solcher Ansätze in den künftigen Diskurs einzuschliessen.

IV. AlgorithmWatch

Mit der sich ständig weiterentwickelnden Technologie geht zunehmende Komplexität einher. Angesichts dessen kann eine effiziente *Algorithm Governance* nicht allein durch die Verabschiedung von Gesetzen sichergestellt werden. Mindestens gleichermassen entscheidend sind Nichtregierungsorganisationen (NGO), welche Ansätze bereitstellen, die die Transparenz und Verständlichkeit von Algorithmen fördern. Der Vortrag von Matthias Spielkamp (Gründer und Geschäftsführer der NGO AlgorithmWatch) zeigt, wie NGO Licht in algorithmische Entscheidungsfindungsprozesse mit gesellschaftlicher Relevanz bringen. Die kritische Auseinandersetzung mit Algorithmen und der Versuch, ihre Funktionalität einer breiten Öffentlichkeit zu erklären, tragen fundamental zum Verständnis von Algorithmen bei. Jüngste Beispiele dafür sind die Medienberichte über die System Risk Indication (SyRI) in den Niederlanden oder die Personalisierung der Google-Suchmaschine. Auch die grösste Kampagne von AlgorithmWatch, OpenSchufa, wirft Licht auf die *Black Box* automatisierter Entscheidungen von einer der bekanntesten deutschen Scoring-Agenturen.

Es gibt jedoch Algorithmen, bei denen Transparenz nicht zu einem optimalen wirtschaftlichen oder sozialen Ergebnis führt. Zu denken ist an Systeme, welche die allgemeine Sicherheit bezwecken oder öffentliche Aufgaben wahrnehmen bzw. unterstützen sollen, beispielsweise wenn ein Algorithmus darüber entscheidet, welche Steuerveranlagungen genauer überprüft oder welche Fluggäste für eine zweite Sicherheitskontrolle herangezogen werden sollen. Spielkamp betont, die laufende Diskussion sollte sich der Unsicherheit widmen, für welche Algorithmen Transparenz erstrebenswert ist und für welche vorzugsweise nicht. Ebenso ist nach wie vor umstritten, ob Transparenz überhaupt einen wirksamen Schutz von Datensubjekten gewährleisten kann.

V. Google AI Principles

Das in letzter Zeit aufgekommene *Machine Learning*, welches sich grundsätzlich von traditionellem Programmieren unterscheidet, hat enorm an Bedeutung gewonnen. Auf diese Unterscheidung sowie auf die damit verbundenen Schwierigkeiten konzentriert sich Dr. Uli Sachs (Senior Product Counsel bei Google) in seiner technischen Einführung.

Die traditionelle Programmierung basiert auf klar definierten Wenn-Dann-Regeln und ist Grundlage für den grössten Teil der Datenverarbeitung. Im Gegensatz dazu werden die Regeln, auf denen *Machine Learning* basiert, aus Daten aufgebaut. Die Maschine lernt bis zu einem gewissen Grad selbst, indem sie die Daten auswertet und daraus Wahrscheinlichkeiten ableitet. Infolgedessen entstehen Regeln erst durch die Aufbereitung einer möglichst grossen Menge an Daten. Gemäss Sachs stellt die Einführung von *Machine Learning* eine grosse Veränderung dar, die unter anderem eine tatsächliche Erkennung von Objekten im realen Leben ermöglicht. Damit gehen neue Herausforderungen und ethische Fragen einher. So auch die Frage, ob – und wenn ja, welche – Maschinen in Zukunft was entscheiden sollen.

---

[3] Julia Angwin et al., ProPublica 23. Mai 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

[4] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung [DSGVO]), ABl. L119 vom 4.5.2016, S. 1–88.

[5] Erwägungsgrund 71 DSGVO.

[6] Art. 35 DSGVO.

[7] Art. 25(1) DSGVO.

[8] Art. 25(2) DSGVO.

Google als global führender Player in KI hat kürzlich sieben Prinzipien für KI-Anwendungen veröffentlicht,[9] die Teil der Präsentation von Daniel Schönberger (Head of Legal bei Google Schweiz und Österreich) sind. Ihm zufolge gibt es nichts Vorherbestimmtes an den Auswirkungen der KI. Gleichzeitig gibt es grosse Möglichkeiten, KI zu nutzen, wie beispielsweise eine sicherere und gerechtere Entscheidungsfindung im Vergleich zu herkömmlichen, menschenbasierten Entscheidungen. Um unerwünschte Ergebnisse zu vermeiden, enthalten die veröffentlichten Grundsätze Regeln darüber, dass KI keine unfaire Verzerrung erzeugen oder verstärken soll. Der erste Grundsatz – die sogenannte Master-Regel – verlangt eine sozial vorteilhafte KI. Die Google-Prinzipien sind im Allgemeinen flexibel konzipiert, um sie künftigen Veränderungen anzupassen.

VI. Wieso und wie Algorithmen und KI zertifizierbar werden sollten

Die Governance-Ansätze für Algorithmen sind vielfältig und reichen von keiner Regulierung über Datenschutzmassnahmen bis hin zur Einrichtung einer Aufsichtsbehörde. Eine weitere mögliche Lösung ist die Zertifizierung von Algorithmen, welche Dr. Houssem Abdellatif (Global Head Autonomous Driving beim TÜV SÜD) fordert, obwohl er damit besondere Schwierigkeiten verbunden sieht.

Ziel der Zertifizierung ist es, Konsumenten vor den negativen Auswirkungen der Technologie zu schützen. Die Tatsache, dass KI nicht erklärbar ist, stellt eine Zertifizierung infrage, die gerade auf die Schaffung von Transparenz abzielt. Eine Lösung hat der TÜV SÜD noch nicht gefunden. Mit «openGENESIS»[10] hat der TÜV SÜD eine offene Kooperationsplattform geschaffen. Mithilfe von Informationen der relevanten Interessengruppen soll schliesslich ein TÜV für KI lanciert werden. Die Festlegung von Sicherheitsanforderungen und anderen notwendigen Kriterien sowie die Entwicklung von Prüfmethoden und -verfahren zur Überprüfung der Sicherheit spielen eine entscheidende Rolle bei der Etablierung einer Zertifizierung für KI.

VII. Sicherheit und Erklärbarkeit der KI

Gemäss Prof. Dr. Jana Koehler (Hochschule Luzern) sind Sicherheit und Erklärbarkeit häufig anzutreffende Schlagwörter im Bereich der KI. Diese, so attestiert sie, sind jedoch mit gewissen Schwierigkeiten verbunden. Koehler legt dar, dass sich zwei Arten von Systemen unterscheiden lassen. Auf der einen Seite stehen die *Learner* wie beispielsweise *Machine* oder *Reinforcement Learning*. *Machine Learning* verallgemeinert Aktionen anhand von vorgegebenen Beispielen und erkennt statistische Muster. *Reinforcement Learning* (verstärkende Lernsysteme) sammelt durch die Ausführung von Aktionen Erfahrungen und passt das Verhalten aufgrund von Feedback an. Auf der anderen Seite erfassen *Solver*, wie beispielsweise Suchalgorithmen, menschliches Fachwissen in formalen Modellen. Solche Systeme haben unterschiedliche Transparenzstufen. Insgesamt ist die Transparenz der KI und damit die Erklärbarkeit gering. Erklärbarkeit zu erreichen, ist laut Koehler äusserst schwierig, manchmal gar unmöglich. Je nach Anwendung und Kontext muss der Erklärungsbedarf definiert werden. Ab einem gewissen Grad ist die Kontrolle über eine Maschine möglicherweise nicht mehr gewährleistet, weshalb sich die Frage aufdrängt, ob eine dauerhafte Kontrolle überhaupt sichergestellt werden muss. Es gibt Ansätze, um das Problem der mangelnden Erklärbarkeit anzugehen, so beispielsweise die «Turing Box».[11] Dieses Programm bewertet intransparente und nicht erklärbare KI, indem es kontrollierte Experimente durchführt und dadurch die *Black Box* der KI enthüllt.

Schliesslich stellt sich die Frage, ob eine erklärbare KI langfristig erstrebenswert ist. Vielmehr könnte die Sicherheit auch dadurch erreicht werden, dass Algorithmen im Voraus getestet und dadurch die Aktionen vorhergesagt werden.

VIII. Panel zu Algorithm Governance

Zum Ende der Konferenz diskutierten die Referierenden, unter der Leitung von Prof. Dr. Isabelle Wildhaber und Prof. Dr. Melinda Lohmann, über mögliche Ansätze einer *Algorithm Governance*.

Im Mittelpunkt der Diskussion stand die Notwendigkeit einer *Algorithm Governance* und die damit verbundenen, vielversprechendsten Lösungsansätze. Alle Referierenden waren sich im Allgemeinen einig, dass in Zukunft eine Form der *Algorithm Governance* erforderlich ist. Die genaue Form konnte allerdings noch nicht definiert werden. Um diese festlegen zu können, muss in erster Linie der Status quo analysiert werden, d.h. welche rechtlichen Instrumente und Institutionen sich bereits heute anbieten und ob sich diese für eine *Algorithm Governance* eignen. Potenzielle Instrumente sind insbesondere Datenschutzmassnahmen, die Zertifizierung beispielsweise durch einen TÜV, Antidiskriminierungsgesetze sowie technische Massnahmen, Kennzeichnungspflichten oder die Förderung einer *Code Literacy*. Mit Sicherheit müssen diese Instrumente für eine bestmögliche *Algorithm Governance* angepasst werden. Ausserdem müssen bestimmte Aspekte wie die Bedeutung und Definition von Algorithmen konkretisiert werden. Meinen wir menschengemachte oder doch eher maschinenbasierte Entscheidungen, wenn wir von Algorithmen sprechen? Eine Analyse des Ist-Zustandes sollte ferner bestimmte bereits geregelte Felder berücksichtigen. Bei Bereichen, die noch nicht geregelt sind, muss sorgfältig überprüft werden, ob eine Regulierung überhaupt notwendig ist. Überregulierung sollte eben-

[9] Sundar Pichai, AI at Google: our principles, 7. Juni 2018, <https://www.blog.google/technology/ai/ai-principles/>.
[10] <https://wiki.eclipse.org/OpenGENESIS_WG>.
[11] <https://turingbox.mit.edu/>.

so vermieden werden wie horizontale Regulierung. Schliesslich geht es darum, sorgfältig zwischen verschiedenen Bereichen und Disziplinen zu unterscheiden und diese nur mit dem geeigneten Instrument zu steuern, wo *Algorithm Governance* erforderlich ist. Es gibt dabei noch viele offene Fragen zu beantworten.

Wie das Ergebnis einer abschliessenden Umfrage zeigte, präferieren überraschenderweise viele Referierende Datenschutzrecht als vorherrschende Lösung, auch wenn dies die Datenschutzexperten nicht für angemessen halten. In einer modernen Welt seien vielmehr systematischere und kollektivere Rechtsbehelfe erforderlich, da die Vorstellung, dass Transparenz, Privatsphäre etc. individuelle Rechte darstellen würden, heutzutage offensichtlich nicht mehr gelte.

IX. Schlussbemerkungen

Diese Tagung ist ein Beitrag zum Austausch über die Notwendigkeit von *Algorithm Governance*. Die Referierenden waren sich einig, dass eine Form von Governance unerlässlich ist. Dennoch gibt es viele ungeklärte Fragen, wie die Gesellschaft mit dem Phänomen «Algorithmus» umgehen will. Die technische Komplexität und der damit zusammenhängende Mangel an Sensibilisierung in der breiten Bevölkerung erweisen sich als Knackpunkt in der Diskussion um geeignete Massnahmen. Das scheinbar unumgängliche Spannungsfeld zwischen Erklärbarkeit und Genauigkeit von Algorithmen führt stets in eine Sackgasse, wenn es darum geht, Transparenz mit maximaler Funktionalität zu verbinden. Daher ist in grundsätzlicher Hinsicht zu hinterfragen, wie mit Algorithmen in einer idealen Gesellschaft umzugehen ist. Das Zusammenführen von Perspektiven verschiedener Disziplinen wie Recht, Informatik und Soziologie spielt schliesslich eine entscheidende Rolle bei der Suche nach und dem Festlegen von geeigneten Ansätzen.