

## Algorithmen fürs Gemeinwohl<sup>1</sup>

Dass die Delegation von Entscheidungen an Maschinen in bestimmten Situationen zu besseren Ergebnissen führt, weiß jeder Autofahrer. Von Zeit zu Zeit fällt immer mal irgendwo eine Ampel aus, so dass Polizisten den Verkehr regeln müssen. Und es ist frappierend, um wie viel schlechter der Verkehr an besagten Kreuzungen fließt – unabhängig davon, welche Personen den Job übernehmen und um welche Straßensituation und Verkehrslage es sich handelt. Immer entsteht bei händischer Verkehrsregelung in irgendeine Richtung ein Kreuzungsstau. Offenbar ist die automatische Lichtzeichenverkehrsregelung dem Menschen also weit überlegen. Unserer Einschätzung nach ist das in vielen gesellschaftlichen Bereichen der Fall. Wir sind überzeugt: Zahlreichen Herausforderungen unserer Zeit können wir nur begegnen, indem wir elaborierte technische Möglichkeiten der Analyse von Daten und des Schlussfolgerns daraus (einschließlich des Vorbereitens und Treffens von Entscheidungen) nutzen.<sup>2</sup>

In einem bis Mitte 2019 laufenden Kooperationsprojekt der Stiftung Neue Verantwortung und der Bertelsmann Stiftung untersuchen wir Möglichkeiten der Förderung des Gemeinwohls durch Datenanalyse- und automatisierte Entscheidungssysteme (im Folgenden kurz: Algorithmen). Wir fragen nach den Bedingungen für die Realisierung des großen Potenzials, das darin liegt, und möchten einen Beitrag zur Versachlichung eines mitunter geradezu hysterischen Gefahrendiskurses leisten.<sup>3</sup> Besonders augenfällig wird diese Prägung durch die deutsche Übersetzung des Buchtitels „Weapons of Math Destruction“ von Cathy O’Neill. Sie lautet: „Angriff der Algorithmen“. Diese Titelgebung verkennt nicht nur, dass Algorithmen – jedenfalls bis auf weiteres – nicht vom Himmel fallen, sondern sie suggeriert auch, dass sie eine grundsätzliche Gefahr für die Menschheit darstellen.

---

<sup>1</sup> Der Autor, Dr. Tobias Knobloch, bedankt sich bei seiner Kollegin Leonie Beining, mit der zusammen das Projekt gestartet wurde, die sich jetzt aber bereits im Mutterschutz befindet, für ihre Unterstützung beim Verfassen dieses Papiers und bei der Konzipierung des Projekts. Ebenso sei namentlich nicht genannten weiteren Kollegen bei der Stiftung Neue Verantwortung und bei der Bertelsmann Stiftung für ihre Anregungen gedankt.

<sup>2</sup> Zu diesen Herausforderungen gehört übrigens die algorithmische Datenverarbeitung selbst. Denn auch die Frage, ob Algorithmen aufgrund der Zugehörigkeit zu einem Geschlecht, einer ethnischen Gruppe, einer Altersklasse oder aufgrund anderer Merkmale diskriminieren, lässt sich aufgrund der großen Datenmengen, um deren Verarbeitung es beispielsweise in Einstellungsverfahren geht, eigentlich nur mit maschineller Hilfe zuverlässig beantworten (und ggf. beheben). Vgl. Venkatasubremian et al. (2015): *Certifying and removing disparate impact*. <https://arxiv.org/pdf/1412.3756v3.pdf>

<sup>3</sup> Eine von wenigen Ausnahmen ist Mohabbat Kar (2017): *Sensoren, Daten, Algorithmen: Regieren nach Zahlen oder das Ende des Politischen*. <https://www.heise.de/ho/meldung/Sensoren-Daten-Algorithmen-Regieren-nach-Zahlen-oder-das-Ende-des-Politischen-3894202.html>

Wir streiten nicht ab, dass es mannigfache Fehlerquellen<sup>4</sup> und handfeste Risiken gibt, die sich in beklagenswerter Form schon realisiert haben. Man denke an Micro-Targeting im Umfeld von Wahlen, an allzu unreflektiert eingesetzte Predictive-Policing-Software, die ohnehin Benachteiligte weiter diskriminiert, an falsch ausgestellte Mahnbescheide, nicht plausible Risiko-Selektionen, unfaire Bewährungsentscheidungen oder aber systematisch im Sozialsystem angelegte, maschinelle Verfestigungen sozialer Disparität.<sup>5</sup> All das wollen wir als Gesellschaft nicht als Ergebnis eines vermehrten Technologieeinsatzes sehen, da dürfte Einigkeit bestehen. Und von anderen, gesellschaftlich enorm einflussreichen und zumindest diskussionswürdigen technischen Phänomenen, wie Online-Werbetechnologie (kurz: Ad Tech) oder High-Frequency-Trading (automatisierter Wertpapierhandel in Hochgeschwindigkeit), haben wir dann noch gar nicht gesprochen. Aber all das spricht aber nicht per se gegen die Automatisierung von Analysen und Entscheidungen, von der wir massiv profitieren können. Und es spricht auch nicht gegen eine intensive Beschäftigung mit diesen Potenzialen und wie wir sie realisieren können, während wir mögliche oder tatsächliche Schattenseiten zunächst systematisch auszublenden versuchen, um ein besseres Gespür für die positiven Effekte zu entwickeln.

### **Chancen der Algorithmisierung (Workshop-Modul 1)**

Das Zusammenwirken von Daten und Algorithmen kann viele Aufgaben, für die Menschen schlussfolgerndes Denken und implizites Können brauchen, fehlerfreier, effizienter und effektiver lösen, darunter zahlreiche Allokationsaufgaben, auch und gerade in der öffentlichen Verwaltung. Bei anderen Herausforderungen, beispielsweise im medizinisch-diagnostischen Bereich, etwa der Krebserkennung und -therapie oder bei der Erkennung und Behandlung seltener Krankheiten, haben sich Computerprogramme als hervorragende Ergänzung menschlichen Wissens und ärztlicher Intuition herausgestellt. Zuletzt gibt es Dinge, die nur Maschinen leisten können, weil unsere kognitiven Kapazitäten bei weitem nicht ausreichen. So ist eine inhaltliche Auswertung und Klassifizierung sämtlicher Forschungsergebnisse, die alleine in der Medizin publiziert werden, nur mit maschineller Unterstützung überhaupt möglich.

Ein Algorithmus knüpft unmittelbar an Zusammenhänge in der physischen Welt an. Um zu funktionieren, braucht er keine von und für Menschen gemachten Gründe - Korrelationen reichen aus.<sup>6</sup> Im Vergleich zur menschlichen Wahrnehmung und Informationsverarbeitung erfüllt er seinen Zweck also mit anderen, eigenen Mitteln. Das ist in der Geschichte der Technikentwicklung ein durchaus altbekanntes Phänomen: Auch Flugzeuge etwa sind nach eigenen Regeln konstruiert worden und nicht mit der

---

<sup>4</sup> Diesbezüglich einschlägig ist die noch unveröffentlichte Studie von Prof. Dr. Katharina Zweig für die Bertelsmann Stiftung (2018): *Wo Maschinen irren können. Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung.*

<sup>5</sup> Vgl. Eubanks (2018): *Automating Inequality. How high-tech tools profile, police, and punish the poor.* <https://us.macmillan.com/automatinginequality/virginiaeubanks/9781466885967/>

<sup>6</sup> Die gesamte Big-Data-Debatte kreist mehr oder weniger um diesen epistemologischen Kern.

Absicht, naturgetreu Vögel nachzuahmen. Überhaupt sind viele Phänomene des Berechenbar-Machens und der Automatisierung, mit denen wir uns heute zu Recht intensiv befassen, weil sie immer mehr an Bedeutung gewinnen, qualitativ nicht wirklich neu. Im Kern handelt es sich um den kybernetischen Traum der Allvermessung von Menschen<sup>7</sup> und Dingen, um sie effizienzgetriebener Steuerung staatlicher, wirtschaftlicher, geheimdienstlicher oder militärischer Art zu unterwerfen.<sup>8</sup> Was tatsächlich neu ist, ist die regelrechte Explosion der Datenmengen, gepaart mit einem linearen Wachstum bei der Rechenkraft. Beides zusammen erfordert immer ausgefeiltere Analysemethoden (die wir hier zusammenfassend als Algorithmen bezeichnen) – zumindest dann, wenn man möchte, dass die Erkenntnisse proportional zur Informationsmenge zunehmen.

Maschinen können Menschen Empfehlungen geben, die auf andere Art nicht zu haben wären. Ich habe nicht genug Freunde und Bekannte, um all die Literatur und all die Musik über persönliche Empfehlungen zu entdecken, die mir Amazon und Spotify schon empfohlen haben. Algorithmen arbeiten konsistent und machen keine Fehler, die Menschen regelmäßig unterlaufen. Das Wertefundament von Algorithmen ist, anders als das von menschlichen Entscheidungsträgern, programmierbar und damit grundsätzlich kontrollierbar. Die Ergebnisse algorithmischer Entscheidungen lassen sich, zumindest im Prinzip, im Nachhinein überprüfen und korrigieren. Durch die Automatisierung von Prozessen werden Menschen von Routinetätigkeiten entlastet, was Freiräume für die Beschäftigung mit drängenderen Problemen und sogar mehr Zeit für menschliche Interaktion (zum Beispiel im Verhältnis zwischen Arzt und Patient oder zwischen Verwaltung und Bürgern) schaffen kann. Angebote und Dienste lassen sich leichter personalisieren oder auf Gruppen zuschneiden. – Dies sind nur einige der tatsächlich beobachtbaren und denkbaren Vorteile.<sup>9</sup> Im **ersten Workshop-Modul** wollen wir solche abstrakten gesellschaftlichen Gewinne, die sich durch den Einsatz von Algorithmen ergeben können, sammeln und in die Kategorien „Effizienzgewinne und Optimierungen“, „Nur maschinell erreichbare Lösungen“, „Verbesserungen des sozialen Miteinanders“, „Stärkung demokratischer Strukturen“ und „Sonstige Chancen“ einzuordnen versuchen.

---

<sup>7</sup> Einen guten Überblick über Scoring-Beispiele bietet Lischka/Klingel (2017): *Wenn Maschinen Menschen bewerten*. <https://algorithmenethik.de/2017/05/02/wenn-maschinen-menschen-bewerten/>

<sup>8</sup> Vgl. Pinto (2017): *Rückkopplungsformen – Behaviorismus, Kybernetik, Autopoiesis*. In: Franke/Hankey/Tuszynski: *Nervöse Systeme*. Berlin: Matthes & Seitz, S.48-65.

<sup>9</sup> Eine Algorithmenentwicklung, die ihren essenziell sozialen Charakter nicht verkennt, hat den weiteren Vorteil, dass im Zuge dessen viele bereits zuvor bestehende Probleme und Schieflagen erkannt und adressiert werden können, weil Datengrundlagen, implizite Annahmen und Vorgehensweisen zum ersten Mal explizit gemacht werden müssen. Vgl. z.B. Schulz & Polster (2017): *Algorithmic decision making and human rights*. <https://www.hiig.de/en/blog/algorithmic-decision-making-human-rights/>

## Positivbeispiele und Kartografierung von Anwendungsfällen (Workshop-Modul 2)

An zwei konkreten Beispielen soll nun das gemeinwohldienliche Potenzial der Nutzung von Algorithmen insbesondere bei der Entscheidungsvorbereitung und Entscheidungsunterstützung verdeutlicht werden.

### Schuleinzugsbereicherechner für Berlin

Die idalab GmbH hat 2016/17 im Rahmen eines Gemeinschaftsprojekts der Technologiestiftung Berlin, der Stiftung Neue Verantwortung, der Berliner Senatsverwaltung für Wirtschaft und Technologie und der Bezirksverwaltung Tempelhof-Schöneberg eine *Intelligent Zoning Engine* für die effizientere und klügere Zuweisung von Grundschulplätzen an schulpflichtige Kinder entwickelt.<sup>10</sup> Diese Aufgabe wurde von der Berliner Verwaltung zuvor mittels Excel-Listen und öffentlich verfügbaren Kartendiensten von Hand wahrgenommen. Das Tool, das mit einem einfach bedienbaren Dashboard ausgestattet ist, erlaubt die automatische Optimierung nach den (gesetzlichen vorgegebenen) Parametern „Schulweglänge“, „Schulwegsicherheit“ und „Schulauslastung“. Daneben sind bei laufender Optimierung auch händische Eingaben (etwa aufgrund von Expertenwissen bei Verwaltungsmitarbeitern) sowie die Berücksichtigung weiterer, weicher Faktoren wie des sozialen Mischungsverhältnisses an einzelnen Schulen möglich, was in einer wachsenden und sich in der Sozialstruktur wandelnden Stadt wie Berlin sehr wichtig ist. Das Besondere an der browserbasierten Anwendung ist, dass sie mit vergleichsweise wenig Entwicklungsaufwand und fast ausschließlich auf der Grundlage von in Berlin bereits als Open Data verfügbaren Datenquellen so weit entwickelt wurde, dass sie sich schon vor der offiziellen Release verwaltungsseitig (inoffiziell) im produktiven Einsatz befindet. Im Workshop wird Dr. Paul von Büнау, Inhaber und Geschäftsführer von idalab, dieses Beispiel zu Beginn des zweiten Moduls vorstellen.

### Schutz von Kindern vor körperlichen Übergriffen

Ein weiteres, in unserer Wahrnehmung grundsätzlich positives Beispiel für das Potenzial von Algorithmen für das Gemeinwohl ist der Einsatz der Prognose-Software *Rapid Safety Feedback* in den USA.<sup>11</sup> Dieses Tool liefert Jugendschutzbehörden eine Vorhersage, welche Kinder mit hoher Wahrscheinlichkeit von häuslichen Gewaltdelikten betroffen sein werden. Als Indikatoren fließen beispielsweise der Empfang von Sozialleistungen, die Elternbiografien, Drogenkontakte, Meldungen von Vorfällen sowie das geografische und soziale Milieu ein. In der Theorie können mit dieser Software knappe personelle Ressourcen der sozialen Arbeit effizienter eingesetzt und dadurch wiederum tatsächliche Schäden von Kindern eher abgewendet werden. In der Praxis zeigt sich allerdings, dass der tatsächliche Wert des Einsatzes von *Rapid Safety Feedback* stark

---

<sup>10</sup> Eine knappe Übersicht inklusive Demoversion ist unter <https://www.intelligent-zoning.de/> verfügbar. Das Tool lässt sich natürlich auch für ähnliche Optimierungsaufgaben einsetzen und ist dafür auch angedacht.

<sup>11</sup> Einen Überblick vermittelt Hurley (2018): *Can an Algorithm Tell When Kids Are in Danger?* <https://mobile.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html>

davon abhängt, welche zusätzlichen Maßnahmen zur Verbesserung des Monitorings und der Betreuung sozialschwacher Haushalte ergriffen werden (z.B. eine Aufstockung des Personals, das auf eine gestiegene Anzahl von Warnhinweisen auch reagieren und diese mit menschlicher Erfahrung/Intuition evaluieren kann. So ist es zu erklären, dass der US-Bundesstaat Florida aufgrund jahrelanger positiver Erfahrungen weiterhin auf den Einsatz der Anwendung zählt, während der US-Bundesstaat Illinois, in dem zwei laut Software aus mit vermeintlich geringem Risiko behafteten Elternhäusern stammende Kinder zu Tode gekommen sind, zukünftig auf den Einsatz des Tools verzichten wird.<sup>12</sup> Das zeigt, dass der soziale Kontext des Einsatzes dieser soziotechnischen Systeme keinesfalls vernachlässigt werden darf, wenn man über ihren gesellschaftlichen Nutzen nachdenkt.

Wir gehen davon aus, dass es zahlreiche weitere tatsächliche oder mögliche Positivbeispiele gibt. Bisweilen ist ein gesellschaftliches Problem auch in einem Land mit maschineller Hilfe bereits gelöst bzw. gelindert worden, dann stellt sich die Frage nach Übertragbarkeit. **Im zweiten Modul** unseres Auftaktworkshops wollen wir Beispiele der oben exemplarisch eingeführten Art sammeln und systematisieren. Wie die Systematisierung erfolgen soll, wird nachfolgend kurz erläutert.

Die oben angeführten Beispiele haben zweierlei gemeinsam: Es handelt sich einerseits um gesellschaftliche Situationen, die wir gerne ändern würden, wenn wir könnten, und andererseits um Problemstellungen, die sich prinzipiell als mathematische Aufgaben reformulieren lassen. Die Kombination dieser Dimensionen ergibt folgende Matrix:



<sup>12</sup> Eine gute Erörterung dieser disparaten Fälle bietet Hustedt (2018): *Die algorithmische Kristallkugel der US-Kinderschutzbehörden*. <https://algorithmenethik.de/2018/01/22/die-algorithmische-kristallkugel-der-us-kinderschutzbehoerden/>

Von größtem Interesse ist für uns das Feld oben rechts, weil hier Problemrelevanz und grundsätzliche maschinelle Lösbarkeit zusammenkommen. In dieses Feld fallen für uns auch die beiden oben eingeführten Beispiele (andernfalls wir sie nicht ausgewählt hätten). Die Einordnung ist zunächst einigermaßen subjektiv. Im Workshop wollen wir sie durch intersubjektive Aushandlung im Expertenkreis zumindest näherungsweise objektivieren. Um weiter sinnvoll mit dieser Matrix arbeiten zu können, braucht man allerdings Klassifikationskriterien. Für den jetzigen Zeitpunkt mag es genügen zu sagen, dass wir weder das Berechenbarkeitspotenzial auf der Y-Achse noch das gesellschaftliche Verbesserungspotenzial auf der X-Achse für unsere Workshop-Zusammenarbeit strikt auslegen wollen. Vielmehr soll es uns für den angestrebten heuristischen Ansatz zunächst ausreichen zu vereinbaren, dass wir jeweils Idealzustände annehmen und die vielfältigen technischen, sozialen und politischen Restriktionen, die in der Realität zweifellos anzunehmen sind, methodologisch ausblenden.

### **Interventionsoptionen für die Chancenrealisierung (Workshop-Modul 3)**

Wichtig ist es in unseren Augen stets zu bedenken, dass es keinen Automatismus gibt, der von guten Absichten und dem Einsatz (vermeintlich) bewährter Technologie zu gesellschaftlich wünschenswerten Ergebnissen führt. Wir müssen in jedem Einzelfall gründlich darüber nachdenken, welches System wir in welchen Zusammenhängen zu welchen Zwecken einsetzen wollen. Und wir müssen das idealerweise tun, bevor die Systeme in der Welt sind, nicht nachdem oder während sie unter realen Bedingungen angewendet werden.<sup>13</sup> Für dieses Nachdenken bezogen auf Einzelfälle bzw. Klassen von Fällen<sup>14</sup> brauchen wir übergeordnete Prinzipien, gemeinsame Kategorien, Methoden und Anknüpfungspunkte, mit denen sich positive Netto-Effekte für die Gesellschaft erzielen lassen. Nach eben diesen Interventionsmöglichkeiten sucht das Projekt „Algorithmen fürs Gemeinwohl“, und diese werden im Auftaktworkshop erstmalig **im dritten Modul** behandelt.

Wir möchten im hier vorliegenden heuristischen Zusammenhang zunächst nicht zwischen einfachen Optimierungs- und Prognose-Algorithmen, Expertensystemen und selbstlernenden (teilautonomen oder vollständig autonomen) Entscheidungssystemen unterscheiden, um nicht von vornherein ggf. besonders potenzialreiche Anwendungen aus dem Blick zu verlieren. Aus diesem Grund schlagen wir für die Analyse möglicher Interventionsoptionen unter Gemeinwohlperspektive bei einzelnen Beispielen und in

---

<sup>13</sup> Eine solche Herangehensweise kann man mit dem Bielefelder Wissenschaftssoziologen Wolfgang Krohn auch als „Realexperiment“ oder „Gesellschaftsexperiment“ bezeichnen. Ein Vorgehen, bei der die Funktionsweise technischer Systeme in vivo an der Gesellschaft getestet wird, halten wir weder politisch noch ethisch für vertretbar. Leider scheint diese Vorgehensweise momentan – sei es bewusst oder unbewusst – in etlichen Bereichen die Regel zu sein.

<sup>14</sup> Natürlich ist nicht jeder Prozess der Entwicklung und des Einsatzes von Software einzigartig, auf der anderen Seite kann man nicht alle über einen Kamm scheren. Vielmehr ist von Familienähnlichkeiten auszugehen.



bestimmten Anwendungsfällen folgende grobe Prozesseinteilung zur Strukturierung der Algorithmenentwicklung vor:

- Abstraktion u. Problemdefinition
- Teambildung
- Modellierung
- Implementierung
- Datensammlung u. -aufbereitung
- Testdatenbetrieb, Evaluierung u. ggf. Überarbeitung
- Überführung in den Produktivbetrieb
- Auditierung u. ggf. Revision
- Überführung in den Regelbetrieb
- Regelmäßige Evaluierung im Regelbetrieb

Daran schließen sich mindestens die folgenden Fragen an: Ist die Darstellung unter Berücksichtigung der oben gemachten Einschränkung korrekt und vollständig? Welche auf das Gemeinwohl einzuhaltenden tatsächlichen Maßnahmen sind in den einzelnen Phasen bekannt, welche sind darüber hinaus denkbar? Was sind die bekannten Hürden für die Berücksichtigung von Gemeinwohlaspekten in den einzelnen Entwicklungsphasen? Welche Akteure und Institutionen sind entscheidend, wenn es darum geht, Maßnahmen tatsächlich zu ergreifen und Hürden zu überwinden? Wie würde in einer idealen Welt in jeder der Phasen eine maximale Berücksichtigung der Gemeinwohlperspektive aussehen?

### **Fazit**

Wir plädieren für einen unaufgeregten, arbeitsteilig organisierten Diskurs über die Möglichkeiten und Grenzen des Einsatzes von Algorithmen. Diese Nüchternheit ist nicht zuletzt deshalb angebracht, weil es letzten Endes um nicht mehr und nicht weniger geht als um die Frage, wie wir im sogenannten Datenzeitalter eine lebenswerte Gesellschaft sein und bleiben können. In diesem Zusammenhang ist es zweifellos angebracht, offenkundige Schwächen und Schwierigkeiten zu untersuchen und möglichst zu beheben. Ebenso notwendig ist es in unseren Augen allerdings die Frage zu stellen, was durch den verstärkten Algorithmeninsatz neu möglich ist, was vorher nicht möglich war. Verschärft könnte man auch sagen: Abgesehen vom Einsatz guter Software, die so programmiert ist, dass sie laufend dazu lernt, haben wir überhaupt keine Chance mehr, aus den gewaltigen Datenmassen, die uns heute zur Verfügung stehen, als Gesellschaft angemessen Kapital zu schlagen. Am größten scheinen die Chancen durch eine Kombination aus Datafizierung und Algorithmisierung aus der Perspektive der Gemeinwohlförderung momentan überall dort zu sein, wo Algorithmen erstens bewusst, ihre möglichen Stärken und Schwächen reflektierend eingesetzt werden, und wo sie zweitens nicht alleine Entscheidungen treffen, sondern menschliche Entscheidungen vorbereiten, unterstützen oder evaluieren.