

„Diktatur der Daten“ Korrelationen und Wahrscheinlichkeit

Von **Viktor Mayer-Schönberger**

Seit Beginn des Menschseins haben wir die Welt beobachtet, um daraus zu lernen, in der Hoffnung, wir könnten diese Welt, in der wir leben, damit ein wenig mehr verstehen. Das, was uns umgibt, mehr und mehr durchschaubar zu machen und damit unserem Menschsein Bedeutung zu geben, unsere Welt zu verstehen, das ist für uns Menschen beinahe so wichtig wie die Suche nach dem Sinn unserer Existenz selbst.

Denn eine Welt, die man nicht versteht, verleitet dazu, sich an Hoffnungen und Vermutungen festzuklammern, auch wenn diese schrecklich vereinfachend sind: dass Zeit zyklisch ist, die Welt flach und wir das Zentrum des Universums sind, das Gott schuf. Und dass, selbst wenn der Tod nicht das Ende ist, wir keine Kontrolle über unser Leben haben.

„Der Wunsch nach Erkenntnis“

Manche in vergangenen Zivilisationen haben diese Vereinfachungen infrage gestellt, und mussten nicht selten dafür mit ihrem Leben bezahlen. Aber ihre Ideen haben überlebt. Und der Wunsch nach Erkenntnis. Jahrtausende lang lebten wir diesen Wunsch, indem wir Erlebtes und Erdachtes in einen größeren Zusammenhang zu stellen trachteten. Wir verbinden damit nicht zuletzt die Reformation, und die Erkenntnis, dass Zeit linear und das diesseitige Leben nicht vorgegeben, sondern für uns Menschen gestaltbar ist.

Im 17. Jahrhundert begannen die Menschen systematisch die Welt als etwas systematisch Erforschbares zu begreifen; sie sammelten Daten und analysierten sie. Dafür entwickelten sie immer bessere Messgeräte und Messmethoden. Die Wirklichkeit wurde, jedenfalls in immer mehr Dimensionen, quantifizierbar. Die "wissenschaftliche Methode" begann so ihren Siegeszug.

Aber das Sammeln und Auswerten von Daten blieb ungemein zeitaufwendig und kostenintensiv. Und deshalb sammelten die Menschen nur genauso viele Daten wie für die Beantwortung einer konkreten bereits gestellten Frage absolut notwendig war. Noch heute sind alle Institutionen, Verfahren und Methoden darauf gerichtet, so wenig Daten wie möglich zu sammeln und zu analysieren. Es ist die Zeit von "small data" - einer zunehmend empirischen Sozialforschung, die von methodischen Abkürzungen wie dem randomisierten Sample profitierte: wieder eine Möglichkeit aus wenig Daten (einem kleinen Sample) auf das Große schließen zu können, ohne das große Ganze selbst gemessen und analysiert zu haben.

All das - diese gesamte Beschränkung auf wenige Daten aber war nur temporär, ein Artefakt des analogen Zeitalters. Denn in der digitalen Zeit sammeln wir immer mehr Daten. Es begann vielleicht vor ein, zwei Jahrzehnten in den Naturwissenschaften. So

konnte mit dem Sloan Digital Sky Survey - einem Teleskop, das im Jahr 2000 in Einsatz ging - in den ersten Wochen seines Betriebs bereits mehr astronomische Daten gesammelt werden als in der gesamten Geschichte der Astronomie.

Auch in der Genetik sind die Fortschritte dramatisch. Im Jahr 2003 ging ein Forschungsprojekt zu Ende, das ein Jahrzehnt lief und eine Milliarde Dollar kostete, und am Ende das komplette Genom eines einzigen Menschen sequenziert hatte. Heute lassen sich diese drei Milliarden Basispaare für ein paar Tausend Dollar in zwei bis drei Tagen sequenzieren.

Aber auch die Internet-Unternehmen sammeln Unmengen von Daten, so dass die gesamte Datenmenge in der Welt heute hunderte Male mehr ist als noch vor einem Vierteljahrhundert – und nahezu ausschließlich digital.

„Die Daten sprechen lassen“

Diese Einfachheit im Sammeln, Speichern, aber auch im Verarbeiten und Analysieren von Daten führt zu einer neuen Qualität der Erkenntnis. Drei Qualitäten charakterisieren Big Data:

Erstens die Möglichkeit, sehr viel mehr und immer öfter nahezu alle Daten eines bestimmten Phänomens zu analysieren als in der Vergangenheit, in der wir uns oft mit Stichproben begnügt haben. Das aber erlaubt uns in der Analyse nicht nur ins Detail zu gehen, sondern auch neue Fragen zu beantworten, die uns vor dem Sammeln der Daten noch gar nicht bewusst waren. Insofern können wir „die Daten sprechen lassen“.

Zweitens gleicht die enorme Menge an Daten auch manches von Qualitätsunterschieden in den Daten wieder aus. Mussten wir in der Vergangenheit sehr viel Energie darauf verwenden, die Stichproben mit der höchstmöglichen Qualität zu sammeln, können wir bei wesentlich mehr Daten hier etwas an Unschärfe zulassen. Die Menge gleicht diese Unschärfe wieder aus.

Drittens, und vielleicht am Wichtigsten, werden mit „Big Data“ Erkenntnisse oftmals schon aus Korrelationen, also scheinbar inneren Zusammenhängen der Daten gewonnen, und nicht erst aus Behauptungen zu Ursache und Wirkung. Big Data kann uns selten sagen „warum“, sondern oftmals nur Auskunft über das „Was“ der Wirklichkeit geben, aber das alleine gibt uns schon wertvolle Einsichten. Denn bei genauerer Analyse müssen wir erkennen, dass eine große Zahl an Ursache-Wirkungs-Zusammenhängen, also die Kausalitäten, auf die wir so stolz sind, sich nicht beweisen lassen, und vielmehr auf (Vor)urteilen beruhen.

Die Korrelationsanalyse durch Big Data eröffnet hier einen demütigeren Zugang zum Verständnis der Wirklichkeit, die eben – so beginnen wir zu verstehen – viel komplexer ist als wir zunächst glaubten.

Ein gutes Beispiel von Big Data bieten die Forschungsarbeiten von Dr. Carolyn McGregor. Sie widmete sich am Universitätsspital in Toronto Frühgeborenen. Diese sind besonders anfällig für Infektionen – und wenn wir Symptome einer Infektion erkennen, ist es für eine erfolgreiche Behandlung mitunter zu spät. Dr. McGregor stellte dieser konventionellen Sicht die Big-Data-Analyse gegenüber: Sie sammelte über digitale Sensoren über 1.200 Einzeldaten der Vitalfunktionen von Frühgeborenen pro Sekunde. Diese große Datenmenge enthielt Daten unterschiedlicher Qualität. In ihnen suchte Dr. McGregor nach Mustern (also Korrelationen), die mit hoher Wahrscheinlichkeit einer Infektion vorhergehen – noch vor dem Auftreten erster Symptome. Sie war erfolgreich – und kann heute mit großer Wahrscheinlichkeit Infektionen ganze 24 Stunden früher vorhersagen, und damit Frühgeborenen das Leben retten. Ein Mehr an Daten, selbst wenn diese unscharf sind, die Auskunft über das „Was“ (nicht aber das „Warum“) geben, sind ausreichend.

„Der wirtschaftliche Wert“

Voraussetzungen für diese und viele andere Big-Data-Erkenntnisse ist unsere Fähigkeit, immer mehr Dimensionen unserer Wirklichkeit in Daten zu fassen und abzubilden, sozusagen zu „datafizieren“. In den letzten Jahren etwa haben wir den Ort datafiziert, nicht zuletzt dank Positionssystemen wie GPS. Aber es geht noch viel weiter: über Facebook wurden Freundschaften datafiziert, über LinkedIn unser Arbeitsumfeld und über Google unsere schriftliche Kommunikation und das, was wir im Internet täglich suchen. Sobald Dinge datafiziert sind, lässt sich aus ihnen Wert schöpfen.

Das bringt uns zum wirtschaftlichen Wert von Big Data. Dieser liegt in der Wiederverwendung der Daten. Bisher wurden Daten in der Regel für einen ganz bestimmten Zweck gesammelt und analysiert, und dann wieder vergessen. Zu teuer war deren Aufbewahrung. In der Zukunft hingegen werden wir erkennen, dass diese Daten für andere Zwecke wiederverwendet werden können und sich daraus nicht nur neue Erkenntnisse, sondern wirtschaftlicher Mehrwert gewinnen lässt.

So verwendet der Paketzusteller UPS die Sensordaten seiner Zustellfahrzeuge, um aus den kleinsten Abweichungen vorherzusagen, wann ein Teil des Fahrzeuges kaputt geht, noch bevor das Teil tatsächlich bricht. Das erlaubt kostensparende vorausschauende Reparaturen, wenn das Fahrzeug ohnehin in der Garage steht.

Die Flugzeuge der deutschen Lufthansa sammelten viele Jahre lang Wetterdaten für den Autopiloten, die dann wieder weggeworfen wurden. Seit ein paar Jahren sammelt die Lufthansa diese Daten und gibt sie dem deutschen Wetterdienst weiter. Allein die

Wiederverwendung dieser Daten hat die Wettervorhersagen in Deutschland um sieben Prozent verbessert.

In den USA gibt es bereits eine ganze Reihe von Startup-Unternehmen, die sich Big Data zunutze machen.

- So konnte das Unternehmen Decide.com die Entwicklung der Einzelpreise von tausenden Konsumgütern überraschend genau vorhersagen, und damit den Konsumenten den optimalen Zeitpunkt zum Kauf geben. Das Unternehmen wurde im Herbst 2013 von eBay gekauft.
- Forscher am MIT haben aus der Veränderung von Preisen in Online-Geschäften die Inflationsrate errechnet – praktisch in Echtzeit – und daraus ein Unternehmen gegründet. Auch das ist eine Wiederverwendung von (Preis)Daten.
- Der amerikanische Discounter Target wiederum kann Verkaufsdaten wiederverwenden und daraus errechnen, ob eine Kundin schwanger ist und wann genau sie ihr Baby erwartet.

Gerade dieses Beispiel deutet freilich auf die bedeutenden Schattenseiten von Big Data. Die offensichtliche ist die mit dem Sammeln, Speichern und Analysieren riesiger Datenberge einhergehende potenzielle Überwachung und der Verlust der Privatsphäre. Denn durch die Wiederverwendung von Daten – einem Kern von Big Data – wird der zentrale Mechanismus des heutigen Datenschutzes ausgehebelt. Heute hoffen wir in der Regel, durch die Information des Betroffenen zum Zeitpunkt des Datensammelns über deren Verwendung und die Notwendigkeit der Zustimmung des Betroffenen Vertrauen zu schaffen und Privatsphäre zu schützen. Aber bei Big Data ist zum Zeitpunkt der Sammlung ja noch gar nicht klar, zu welchen Zwecken Daten wiederverwendet werden. Das Ergebnis sind schwammige Zustimmungformeln, die dem Geist des Datenschutzes widersprechen.

Die zunehmende Wirkungslosigkeit des Datenschutzes ist freilich nur eine relativ offensichtliche Schattenseite von Big Data. Viel weniger bekannt, aber mindestens so problematisch ist die Gefahr, dass Big Data-Vorhersagen missbraucht werden, um Menschen zur Verantwortung zu ziehen, nicht für das, was sie schon getan haben, sondern was die Big Data Analyse lediglich vorhersagt, das sie tun werden. Wer jetzt an den Hollywood-Film „Minority Report“ denkt, der ist auf der richtigen Fährte.

Schon heute werden in vielen Bundesstaaten der USA bei der Entscheidung, ob ein Häftling auf Bewährung freikommt, Big Data Analysen eingesetzt, die vorhersagen wollen, mit welcher Wahrscheinlichkeit der Häftling nach Freilassung in einen Mord verwickelt sein wird. Und in einer immer größer werdenden Zahl an amerikanischen Städten nutzt die Polizei Big Data um vorherzusagen, wann und wo die nächsten Verbrechen geschehen werden, um damit ihre Ressourcen-Planung zu optimieren.

Aber die Gefahr einer Bestrafung aufgrund von Vorhersagen ist nicht bloß auf Polizei und Justiz beschränkt. US-Versicherungen geben Musterschülern einen Rabatt auf ihre Autoversicherung, weil Korrelationen zeigen, dass diese weniger Unfälle verursachen.

Durch Big Data Analysen werden immer öfter Menschen aufgrund von Vorhersagen ihres individuellen Verhaltens anders, und damit eben auch schlechter behandelt werden. Vielleicht erkennt in Zukunft ein Onlinehändler, dass wir ein Produkt wirklich wollen und bietet es uns deshalb zu einem höheren Preis an. Oder wir errechnen, welches Verhalten wem nicht gut tut und üben dann direkten gesellschaftlichen Druck auf die Person aus, ihr Verhalten anzupassen, lediglich auf der Basis von probabilistischen Vorhersagen.

Im Kern verweigern wir den Menschen damit ihren freien Willen, indem wir sie bestrafen noch bevor sie ein Verhalten gesetzt haben. Sie haben keine Chance mehr, ihre Unschuld zu beweisen und keine Möglichkeit mehr, sich in der Situation gegen das vorhergesagte Verhalten zu entscheiden. Wir sagen ihnen, dass sie ohnehin „schuldig“ sind, weil die Big Data Analyse das errechnet hat.

Das wäre das Ende einer ganz zentralen Freiheit des Menschen und der Beginn einer schrecklichen Knechtschaft, gegen die eine Orwell'sche Überwachung nachgerade harmlos erscheinen mag.

Das Problem freilich ist nicht die Big Data Analyse an sich, sondern wie wir Menschen die daraus ermittelten Ergebnisse verwenden. Das Problem entsteht, wenn wir Big-Data-Korrelationen, die nichts über Verursachung aussagen, missbrauchen, um damit über individuelle Verantwortung und Schuld zu sprechen. Freilich: Wir Menschen sind konditioniert, Ursachen zu „sehen“, auch wenn wir ganz falsch liegen. Das gibt uns ein Gefühl, die Wirklichkeit zu verstehen, auch wenn wir uns täuschen. Und das bedeutet, dass wir Menschen eben stets in Versuchung sind, Big-Data-Analysen so zu missbrauchen.

Allgemeiner formuliert laufen wir ständig Gefahr, den Datenanalysen mehr Bedeutung zuzuschreiben als diese verdienen. Wir begeben uns damit in eine Diktatur der Daten, und geben unser kritisches Urteilsvermögen ab.

Das Big-Data-Zeitalter bedarf daher einer klaren gesellschaftlichen (und damit rechtlichen) Regulierung. Die Datenverwendung muss im Sinne eines effektiven



Datenschutzes in Zukunft reguliert werden. Der menschliche freie Wille muss grundrechtlich garantiert werden, genauso wie bisher etwa die Verfahrensgerechtigkeit. Und wir müssen sicherstellen, dass die Daten nicht alle in den Händen einer kleinen Zahl an Quasi-Monopolisten landen, denen dann alle informationelle Macht zukommt. Die Politik ist also umfassend und unmittelbar gefordert.

„Ein Schatten der Wirklichkeit“

Big Data kann uns helfen, unsere Welt besser zu verstehen. Mit Big Data werden wir besser Entscheidungen treffen, welche medizinische Behandlung für uns geeigneter ist, wie unsere Kinder effektiver lernen können und wie Autos sich selbst fahren.

Aber Big Data bedeutet auch, auf der Hut zu sein vor dramatischen Herausforderungen. Wir müssen sicher stellen, dass wir Menschen weiterhin die Herrschaft über diese Technologie behalten und nicht deren Untertan werden.

So viel wir auch aus den Daten gewinnen können, und so wichtig es ist, aus den Daten zu lernen, so entscheidend ist es daher, einen Platz für uns Menschen zu bewahren – für unsere Originalität, für unsere Irrationalität und für die Freiheit, uns auch bewusst gegen die Datenanalyse zu entscheiden. Denn die Daten sind am Ende immer nur ein Schatten der Wirklichkeit und damit immer unvollständig. Wir sollten dieses mächtige Werkzeug daher vor allem mit einem hohen Maß an Demut einsetzen – und mit ebenso viel Menschlichkeit.

Viktor Mayer-Schönberger ist Professor for Internet Governance and Regulation an der Oxford University.

Anmerkung:

Diese Fassung der Rede von Viktor Mayer-Schönberger entspricht der Veröffentlichung in epd medien vom 30. Mai 2014.