

### Big Data

*Big Data*, die Auswertung großer Datenmengen durch immer schnellere Computer, ist das neueste Versprechen der Allwissenheit. Nutzer\*innen dieser Technologie wird versprochen, dass sie nicht mehr mit Stichproben arbeiten müssen, sondern alles über alle sehen können. Händler\*innen nutzen Big Data, um Interesse an Produkten vorherzusagen. Google zeigt die Suchergebnisse an, die die Person vor dem Bildschirm interessieren könnten. Und Geheimdienste versprechen Sicherheit vor Anschlägen, wenn sie nur den gesamten Verkehr im Internet belauschen könnten.

#### Die Technologie

Hinter Big Data stecken nicht nur schnellere Rechner und größere Speicher, sondern auch Fortschritte in Mathematik und Informatik, die viele Fragestellungen überhaupt für maschinelle Analysen zugänglich machen.

Während klassische Verfahren auf objektive Aussagen zielen, operieren moderne Analyseverfahren mit Ungewissheiten und nutzen subjektive Einschätzungen von Expert\*innen. Datenanalyse wird zum automatisierbaren Lernprozess, in dem mehr Daten die Ungewissheit verkleinern, aber nicht zum Verschwinden bringen.

Traditionell waren Datenbestände voneinander isoliert. Banken wussten über Zahlungen, Supermarktketten über Einkäufe und Krankenkassen über Behandlungen. Das Zusammenführen dieser Daten wurde durch technische Schranken und semantische Unterschiede behindert. In der neuen Welt von Big Data machen Standards es leichter, bisher getrennte Datenbestände miteinander zu verknüpfen. Die Idee eines digitalen Zwillinges von Dingen und Menschen rückt technisch zunehmend in den Bereich des Möglichen.

Der neueste Trend sind cyber-physikalische Systeme, in denen Informationstechnologie mit Geräten in der realen Welt oft über weite Entfernungen verbunden ist. Eine zentrale Steuerung kann Daten mittels Sensoren sammeln und auf Gegenstände in der realen Welt einwirken. Beispielsweise kursieren Ideen, mittels vernetzter Fitnessarmbänder und anderer am Körper befestigter Sensoren Empfehlungen für ein individuelles Fitness-Programm zu geben. Es braucht wenig Fantasie, sich vorzustellen, dass Sanktionen oder Bonusprogramme von der Einhaltung dieser Vorgaben abhängig gemacht werden könnten.

## Daten und Geschäft

Ein sichtbares Anwendungsgebiet von Big Data ist Werbung. In der Vergangenheit wurde sie auf Marktsegmente zugeschnitten, in die Menschen mittels Marktforschung eingeteilt wurden. Durch Big Data konnten die seit den 1990er Jahren wirkungsloser werdenden Instrumente des Massmarketings durch direkte Ansprache potenzieller Kund\*innen ergänzt oder ersetzt werden.

Auch bei datenbasierter Werbung sind Fehler mit unerfreulichen Ergebnissen für die Werbetreibenden nicht selten. Wenn eine 16-jährige aufgrund eines geänderten Kaufverhaltens bei Pflegeprodukten von der Supermarktkette Werbung für Babywäsche bekommt, ist ein Anruf des empörten Vaters noch das kleinste Problem. Reputationsschäden, Aufwand für nutzlose Kontakte und Rückschläge im Kampf um das knappe Gut Aufmerksamkeit treiben Fehlerkosten in die Höhe.

Dies wird neben dem Sammeln von immer mehr Daten mit der Nutzung sozialwissenschaftlicher Methoden wie der Ethnographie und der Übernahme naturwissenschaftlicher Ansätze unter Begriffen wie „social physics“ beantwortet. Daneben werden immer ausgefeiltere Methoden zur Beobachtung von Menschen entwickelt, ermöglicht unter anderem durch zunehmende Kontrolle der Vermarkter über mit dem Internet verbundene Geräte ihrer Kund\*innen wie Mobiltelefone, Notebooks oder Smart-TVs. Durch offene und verdeckte technische Mechanismen soll gewährleistet werden, dass Daten korrekt erhoben und zugeordnet werden.

Die so gewonnenen Daten sind auch Handelsware. Auf Basis von Daten individualisierte Werbung, Rohdaten zur Verbesserung von digitalen Zwillingen und vollständige Profile können käuflich erworben werden. Die Möglichkeit von Voraussagen mit einer für ein Geschäftsmodell ausreichenden Genauigkeit wird zur Entwicklung und Vermarktung von Dienstleistungen genutzt. Dies reicht vom vernetzten PKW, der vorsorglich einen Werkstatttermin vereinbart, über den digitalen Assistenten, der einen Tisch für ein Dinner zu zweit bestellt, bis hin zu Filterung von Nachrichten aufgrund vermuteter Vorlieben der Benutzer\*innen.

Auch wenn der Anspruch erhoben wird, sich an den Menschen anzupassen, passen sich die Systeme in Wirklichkeit an den digitalen Zwilling und seine Präferenzen an. Vorhersagen und darauf aufbauende Interaktionen verändern die Wirklichkeit des Menschen und drängen sein Verhalten in die Richtung seines digitalen Zwillingen. Als nächster Schritt folgt die Digitalisierung ganzer Gemeinschaften durch Projekte wie *Smart Home* und *Smart*

*City* und damit möglicherweise auch die Anpassung ganzer Gemeinschaften an ihre digitalen Abbilder.

## Big Data und der Staat

Die Ordnung von Gesellschaften, um die zugewiesenen Aufgaben wie soziale Absicherung, Infrastruktur sowie innere und äußere Sicherheit überhaupt erledigen zu können, ist ein altes Arbeitsprinzip bürokratischer staatlicher Strukturen. Dabei wird auf zentrales, monopolisiertes Wissen und eine möglichst gute Sichtbarkeit aller Aspekte menschlichen Lebens gesetzt.

Daher ist es nicht verwunderlich, dass staatliche Stellen auf den Zug „Big Data“ aufgesprungen sind. Ein naheliegendes Beispiel sind Planungen, z.B. die Auslegung von Verkehrsinfrastrukturen oder Stromleitungen. In der Entwicklungszusammenarbeit wird die Auswertung von Mobiltelefonnutzung oder Satellitenbildern erprobt, um festzustellen, wo Interventionen eine möglichst gute Zielerreichung ermöglichen. Die staatliche Vision von *Smart City* ist ein algorithmisch gesteuertes, zentral geplantes Lebensumfeld.

Der Gefahr, dass viele staatliche Interventionen zwar in bester Absicht geschehen, aber mangels Einbeziehung von Betroffenen und deren Wissen nicht zum Erfolg führen, versucht man wie in der Wirtschaft durch mehr und genauere Daten und Modelle Herr zu werden. Das Big Data innewohnende Versprechen der Sichtbarkeit aller Individuen bedeutet auch die Verlockung, nur mühevoll kontrollierbares, lokales Wissen und Mitgestaltungsansprüche von eigensinnigen Bürger\*innen überflüssig zu machen.

Zu den eifrigsten Datensammlern gehören Sicherheitsbehörden. Aktivitäten in Mobilfunknetzen und sozialen Netzwerken sowie demographischen Daten sollen Orte von Verbrechen oder Unruhen voraussagen. In einer Reihe von US-Bundesstaaten ziehen Richter\*innen maschinell erzeugte Profile von Angeklagten heran, um vorherzusagen, ob eine Wiederholungstat zu erwarten ist, und dies bei der Bestimmung des Strafmaßes zu berücksichtigen. Dass die Algorithmen dabei einen rassischen Bias haben, ist nicht verwunderlich, da sie mit Vergangenheitsdaten kalibriert werden und somit nicht zuletzt die Vorurteile von Sicherheitsbehörden und Gerichten widerspiegeln. Hinzu kommt, dass alleine die Auswahl, welche Daten – z.B. Ethnizität oder Geschlecht – überhaupt analysiert werden, die Schlüsse des Algorithmus schon in eine Richtung lenkt.

Die in den Daten abgebildete Vergangenheit wird damit gleichzeitig zur Voraussage künftiger Ereignisse. Will man Veränderungen in Betracht ziehen, gibt es zwar gute Verfahren, bestehende Trends fortzuschreiben,

sprunghafte, chaotische Entwicklungen entziehen sich allerdings den Big-Data-Analysen.

Das Internet und insbesondere die sozialen Netzwerke machen mittlerweile einen großen Teil der medialen Öffentlichkeit aus. Sowohl Internet-Firmen als auch staatliche Stellen haben ein Interesse daran, dass als verbotene oder als schädlich angesehene Inhalte dort nicht mehr sichtbar sind. In von diesen Firmen und einigen autoritären Staaten aufgebauten Filtersystemen entscheiden „content moderators“ über durch Algorithmen gemachte Vorschläge, Inhalte zu löschen. Das Ergebnis dieser Entscheidung geht auch in die digitalen Zwillinge der Personen ein, die die Inhalte eingestellt haben. Die Nutzung sozialer Medien bei der Organisation sozialer Bewegungen etwa im arabischen Frühling oder beim Sturz von Blaise Compaoré in Burkina Faso im Jahr 2014 haben zu einem zunehmenden Interesse an derartigen Filtertechnologien bei autoritären Staaten speziell in Afrika geführt, wobei hier meist auf chinesische Beratung und Technologie zurückgegriffen wird.

Die Qualität der Modelle und damit letztlich die Bedeutung der genutzten Daten ist oft fragwürdig. Bei Sicherheitsapparaten wie Militär und Geheimdiensten scheint dies kein ernstzunehmendes Hindernis für die Verwendung dieser Technologie zu sein. Das Risiko schwerwiegender Konsequenzen für einzelne Menschen wird als vernachlässigbar gegenüber der Vermeidung schweren Unheils, etwa eines terroristischen Anschlags, angesehen. Im Zentrum der Strategie der USA im Kampf gegen den Terror stehen mit den Drohnen die ausführende Teile eines cyber-physikalischen Systems, dessen Zielauswahl zu einem guten Teil auf den weltumspannenden Datensammlungen der NSA basiert. Die beim Auswerten mit Big-Data-Verfahren zwangsläufigen Fehler haben dazu beigetragen, dass nicht nur Mitglieder der bekämpften Organisationen getötet, sondern auch Hochzeitsgesellschaften oder andere Unbeteiligte bombardiert wurden.

## Folgen und Reaktionen

In der Auseinandersetzung mit den Folgen dieser Technologie wird oft betont, dass Daten über ein Individuum unter dessen Verfügungsgewalt stehen sollten. Die konsequente Einhaltung des Prinzips, dass Staaten und Unternehmen nur über die für die Erledigung ihrer Aufgaben absolut notwendigen Daten verfügen dürfen, sollte helfen, für Menschen nachteilige Konsequenzen von Big Data zu verhindern. Je komplexer die in den Daten abgebildeten Beziehungen sind, desto schwieriger ist jedoch die Zuordnung zu einer Eigentümer\*in. Zumindest ein Teil des Geflechtes an Daten

sollte daher als Gemeineigentum angesehen und allen Menschen verfügbar gemacht werden.

Darüber hinaus muss für Bürger\*innen, Kund\*innen, ... erkennbar sein, welche Daten über sie gesammelt werden. Dabei geht es nicht nur um Rohdaten, sondern auch um die Daten, die bei Analysen entstehen. Es muss auch nachvollziehbar sein, mit welchen Modellen gearbeitet wird, um Wissen zu gewinnen und Vorhersagen zu machen. Individuen werden aufgrund von Komplexität und großen Datenmengen eine solche Offenheit nicht immer für sich nutzen können. Jedoch ermöglicht die Transparenz des Big-Data-Dschungels externe Kontrolle, erleichtert die Reaktion auf Entscheidungen von Algorithmen und hilft, über Datenjournalismus und wissenschaftliche Analysen unabhängiges Wissen zu erzeugen.

Wenn schließlich Big-Data-basierte Algorithmen schwerwiegende Entscheidungen treffen, stellt sich die Frage, nach welchen ethischen Grundsätzen dies geschieht. Zwar ist in extremen Fällen wie vollautonomen Waffen ein Verbot denkbar. Für viele Alltagssituationen fehlen uns aber noch die Antworten, es sei denn, man akzeptiert, dass Alternativen anhand eines errechneten, oft monetären Werts beurteilt werden.

*Erich Pawlik*