



Big Data mache die Forschung obsolet,  
schrieb „Wired“ vor zehn Jahren.



von PHILIPP ALBRECHT

**R**

Rudolf Sommer (74) ist eine skeptische Natur. Der pensionierte Spediteur weiss darum genau, was mit seinen Daten passiert. Die Werbung für seinen Briefkasten steuert er zum Beispiel selber. «Ich will keine Prospekte mit Hörgeräten und Lesebrillen», sagt der Aargauer. «Ich weiss selber, dass ich alt bin.» Von der KünzlerBachmann Directmarketing AG in St. Gallen, die seine Daten an Markenhersteller und Detailhändler verkauft, wollte er etwa wissen, was sie jahrelang über ihn gesammelt hat. In der Antwort zeigte sich, dass ihn die Datensammler als Single einstufen. Dabei ist Sommer seit bald 50 Jahren verheiratet.

Wie ist das möglich bei einer Firma, die mit 100 Jahren Erfahrung im Datensammeln wirbt und auf der Website ihre «innovativen Lösungen zur Verwaltung grosser Datenmengen (Big Data)» hervorhebt?

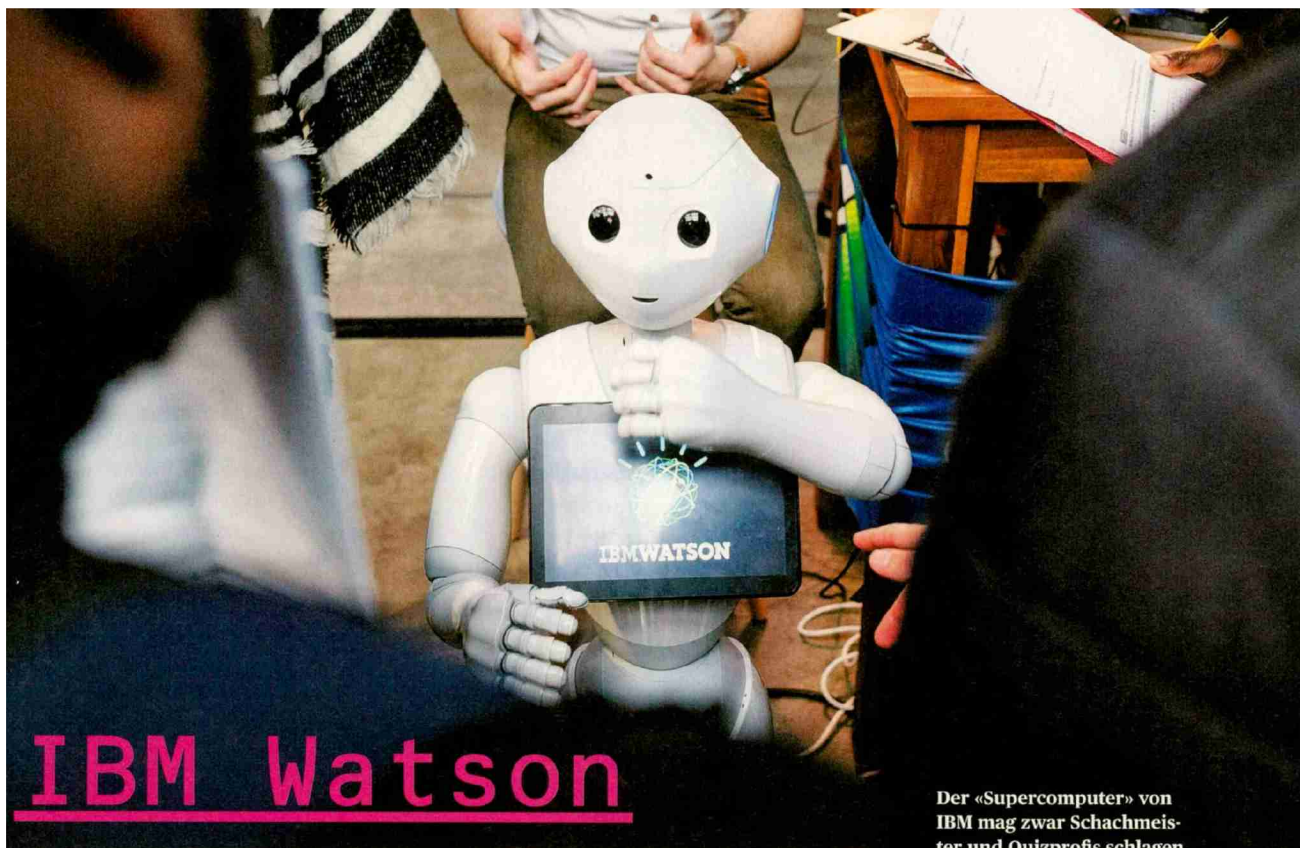
**SAP VERWIRRT DEUTSCHE ELF**

Das könnte man auch SAP fragen. Am Tag des Eröffnungsspiels der Fussball-WM stellte der Softwarekonzern «technologische Innovationen» vor, die der deutschen Nationalmannschaft «zu Spitzenleistungen verhelfen» sollten. SAP Sports One könne «anhand von Mustererkennung und Positionsdaten einfacher und schneller komplexe Spielsituationen und taktische Ausrichtungen der Gegner identifizieren». Die Software soll «unterschiedliche Datenquellen eines Matches nutzen» und so die «Siegchancen optimieren». Bekanntermassen hat es die deutsche Elf nicht über die Vorrunde hinaus geschafft. Es scheint fast so, als hätte die Software die verwirrten Spieler noch zusätzlich überfordert.

Vom grossen Big-Data-Versprechen bleibt nicht viel übrig. Ziemlich genau ▶

▶ zehn Jahre ist es her, seit Chris Anderson, damaliger Chefredaktor des Tech-Magazins «Wired», in einem viel beachteten Artikel schrieb, Big Data werde die herkömmliche Forschung überflüssig machen. Allen voran werde die sogenannte wissenschaftliche Methode nutzlos, die mit einer Fragestellung und viel Recherche beginnt und mit einem Experiment und einer Schlussfolgerung endet.

Auf Andersons Pamphlet folgte die Euphorie. Es machte sich der Konsens breit, dass Wissenschaft und Wirtschaft bald von einem komplett neuen datengetriebenen System beherrscht würden. Die Propheten der Digitalisierung verkündeten einen beschleunigten technologischen Wandel. Neue Geschäftsmodelle sollten dereinst ganze Branchen wegdisruptieren. «Es wird die Weise, wie wir über Gesundheit, Erziehung, Innovation und vieles mehr denken, völlig umkrempeln und Vorhersagen möglich machen, die bisher undenkbar waren», schrieb der österreichische Rechtswissenschaftler Viktor Mayer-Schönberger 2013 in seinem oft zitierten Buch



## IBM Watson

Der «Supercomputer» von IBM mag zwar Schachmeister und Quizprofis schlagen, doch an die Kompetenz von Ärzten kommt er nicht heran. Noch schlimmer: Spitäler brachen die Zusammenarbeit ab, weil Watson mit der Verarbeitung von medizinischem Wissen überfordert war. **Ihr Fazit: mehr Marketing als Maschinenmedizin.**



«Big Data - Die Revolution, die unser Leben verändern wird».

Beratungsfirmen gerieten ob des Hypes in einen Zahlenwahn. McKinsey wollte 2011 ausgerechnet haben, dass Big Data das US-Gesundheitssystem um 300 Milliarden Dollar günstiger machen sollte - jährlich. Es kam dann anders. In vielen Branchen passierte das Gegenteil: In der Medizin verursacht Big Data hohe Kosten und zahlreiche Flops. Etwa in den Spitälern, die blind IBM gefolgt waren. Der IT-Konzern hatte versprochen, dass sein Cloud-basierter Supercomputer Watson Ärzte bei der Diagnose von Krankheiten übertrumpfen werde. Mehrere Einrichtungen sprachen Datensätze und Gelder, um Watson zu füttern. Doch das renommierte Krebszentrum MD Anderson der University of Texas brach 2017 nach drei Jahren und 60 Millionen Dollar an Investitionen das Experiment ab. Auch die Uni-Kliniken Giessen und Marburg beendeten das Watson-Abenteuer frühzeitig, wie der «Spiegel» kürzlich berichtete. Schnell sei klar geworden, dass Watson mehr Marketing als Maschinenmedizin sei. «Ich dachte mir: Wenn wir da weitermachen, investieren wir in eine Las Vegas-Show», gab der oberste Verantwortliche in Marburg zu Protokoll.

«Das Ganze ist ein Bullshit-Generator», bemerkt Gerd Antes (69). Der Medizinstatistiker leitet das Institut für Evidenz in der Medizin in Freiburg im Breisgau und ist wissenschaftlicher Vorstand der Cochrane-Deutschland-Stiftung. Neben Watson hat er in seinem Fachgebiet viele andere Big-Data-Versuche scheitern sehen. Bis heute ist ihm kein einziges Projekt bekannt, das seine hohen Investitionen wert wäre: «Ich sage nicht, dass alles schlecht ist, es fehlt mir einfach der Qualitätsbegriff bei Big Data. Es gibt keine Beweise.»

Sein Paradebeispiel gefloppter Projekte ist Google Flu Trends. Google ist zwar einer der ganz wenigen Konzerne, die mit Big Data tatsächlich Geld verdienen. Doch der

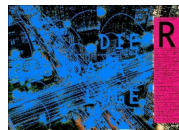
Versuch, eine künstliche Grippe-Früherkennung aufzubauen, scheiterte grandios. Aus den Millionen von Suchanfragen zu Symptomen versuchte Google ab 2008 ein Muster herauszufiltern, um bei der Prognose von Grippewellen schneller zu sein als die herkömmlichen Systeme, die auf Ärzte-Umfragen basieren.

Die ersten zwei Jahre klappte das ganz gut, die Resultate deckten sich zu 97 Prozent mit jenen der alten Messmethode und konnten deutlich früher ermittelt werden. Doch dann brach das System ein. Zwischen 2011 und 2013 sah die Software Grippewellen, wo im Nachhinein keine einzige Arztkonsultation registriert werden konnte. Google kommunizierte nie öffentlich über die Gründe für den Shutdown nach nur vier Jahren. Auf der Homepage schreibt der Konzern, man sei «gespannt, was als Nächstes kommt».

#### UNGEFILTERTE KORRELATIONEN

Ein möglicher Grund fürs abrupte Ende: Die Google-Ingenieure wussten nicht, wie der Rechner zu seinen Resultaten kam. Offenbar waren sie nicht in der Lage, eine Verbindung zwischen den Suchbegriffen und dem Ausbruch der Grippe herzustellen. Was sie hatten, waren statistische Muster: Korrelationen. Hier liegt das Hauptproblem von Big Data. Es gibt Korrelationen zuhauf, aber kaum bis gar keine Kausalität. Für Flu Trends heisst das: Nur weil in Bern plötzlich mehr Menschen Grippe-symptome googeln, muss das nicht bedeuten, dass dort der Virus ausgebrochen ist. Sie könnten auch einfach im Fernsehen einen Beitrag über eine schwere Grippewelle im Südtirol gesehen und sich dann im Internet über Symptome informiert haben.

Kausalitätsbezüge herzustellen, ist gefährlich. Nur weil sich Faktor A und Faktor B parallel verändern, heisst das nicht, dass sie zusammenhängen. Statistiker haben über Jahrhunderte gelernt, falsche Korrelationen zu filtern. Algorithmen schaffen



das bis heute nicht. Big-Data-Skeptiker machen sich einen Spass daraus. Die Website Tylervigen.com hat Grafiken übereinandergelegt, die sich zufällig gleich verhalten, aber nichts miteinander zu tun haben. Eine zeigt die scheinbare Parallelität zwischen der Absatzentwicklung japanischer Autos in den USA und der Anzahl sterbewilliger Menschen, die vor Autos sprangen. Eine andere zeigt, dass der Pro-Kopf-Konsum von Käse genauso stark angestiegen ist wie die Anzahl Menschen, die im Schlaf starben, weil sie sich in ihrer Bettwäsche ver-

hedderten. Der Computerforscher David Bailey umschreibt die Sache so: «Computer, die mit Big Data operieren, produzieren so schnell Unsinn wie nie zuvor.»

Trotz wachsender Flop-Rate gibt es noch überzeugte Verfechter. Marcel Salathé etwa, Professor für digitale Epidemiologie an der ETH Lausanne. Er war enttäuscht über das Ende von Flu Trends, einem Projekt, das er intensiv beobachtet und das «erstaunlich gut funktioniert» habe. «Flu Trends war das Nebenpro- ▶



Der deutsche Softwarekonzern SAP hat für «Die Mannschaft» ein komplexes Big-Data-Produkt entwickelt, das der deutschen Elf an der WM-Endrunde in Russland «zu Spitzenleistungen verhelfen» sollte. Die Innovation, vollgepackt mit Machine-Learning-Tools, brachte aber bekanntermassen nicht die gewünschten Resultate.



**oBike**

Der Fahrradverleiher aus Singapur drang bis Frühling 2018 in 24 Länder vor. Geld verdienen wollte man mit den Bewegungsdaten der Velos, unter anderem via Inserate auf der Bestell-App. Doch die Daten hatten offenbar nicht die nötige Qualität. Im Sommer ging dem Start-up das Geld aus.

Ausser Facebook und Google verdienen nur vereinzelte Firmen Geld mit Big Data.



# „Die Datentöpfe sind voll, müssen aber noch verknüpft werden.“ Herbert Bolliger

▶ jekt eines 20-Prozent-Ingenieurs», ist er überzeugt. «Die Algorithmen wurden einfach nicht mehr aktualisiert, und darum funktionierte es irgendwann nicht mehr.» Dass die Big-Data-Kritiker das Beispiel gerne verwenden, um ihre These zu stützen, sei unfair. «Die Google-Ingenieure sind nicht blöd, das Unternehmen schießt ständig solche Projekte ab, das ist nichts Aussergewöhnliches.»

Ein Grundproblem liegt seiner Meinung nach bei den falschen Versprechen, die im Umgang mit der Datenflut abgegeben wurden. Die frühen Euphoriker und ihre Jünger sahen in Big Data Big Business. Doch auf raffinierte Businessmodelle warte man noch immer: «Da verkauft man die Technologie zu teuer», sagt Salathé. «Es hat bisher noch keiner ein ausserordentlich profitables Big-Data-Businessmodell gefunden, das nicht mit Werbung zu tun hat.»

Facebook und Google haben die Disziplin perfektioniert. Werbung ist ihre grösste Einnahmequelle. Von den 13,2 Milliarden Dollar, die Facebook im zweiten Quartal 2018 umsetzte, kamen 13 Milliarden aus der Werbung. Bei Google macht der Werbeanteil 85 Prozent aus. Abseits der Silicon-Valley-Konzerne gibt es nur vereinzelt Firmen, die mit Big Data Geld verdienen.

Gescheiterte Versuche aber gibt es zuhauf. Etwa den des Fahrradverleihers oBike, der bis vor kurzem Städte in 24 Ländern bediente, darunter Zürich. Die Haupteinnahmequelle sollte nicht die Mietgebühr für die Velos sein, sondern die Werbung, die man auf der Bestell-App zu schalten gedachte. Die gefahrenen Routen hätten Daten liefern sollen, die man auf die Bedürfnisse der Inserenten zuschneiden würde. Doch es biss keiner an. Unter anderem wollte man die Daten Händlerketten verkaufen, damit diese sähen, welche Routen ihre potenziellen Kunden auf oBikes

bevorzugten. Doch die wussten schon, wie die Besucherströme fliessen - ohne Big Data. Im Juli meldete oBike Insolvenz an.

«Es gibt einen grossen Hype mit viel Begeisterung, aber wenn ich mit den Verantwortlichen in den Unternehmen spreche, spüre ich viel Ernüchterung», sagt Christian Westermann, Daten-Experte und Partner bei PwC Schweiz. Er hat zwei Stolpersteine ausgemacht: zu wenig Daten und solche von schlechter Qualität. oBike fiel über den ersten Stein. Wer nicht genug gutes Material hat, kann darauf selbst mit den neuen Methoden des Machine Learnings keine robusten Modelle aufbauen und Schlüsse daraus ziehen. Geschweige denn Werbung verkaufen.

Das heisst aber im Umkehrschluss nicht etwa, dass möglichst viele Daten die besten Resultate bringen. «Das Credo *«Je mehr, desto besser»* geht hier nicht auf», weiss Manuel Nappo, Institutsleiter Digital Business an der Hochschule für Wirtschaft Zürich. «Es hat sich gezeigt, dass es nicht die Menge der Daten macht, sondern deren

Qualität.» Die Computerwissenschaftler Cristian Calude und Giuseppe Longo kamen vor zwei Jahren bei einer Studie sogar zu einem überraschenden Schluss: Zu viel Information hat ähnliche Auswirkungen wie zu wenig. Mit zunehmender Datenmenge häufe sich die Anzahl Korrelationen, was die Gefahr von willkürlichen Resultaten erhöhe.

## DATEN BRAUCHEN INFOS

Um an Qualitätsdaten zu gelangen, muss man einigen Aufwand betreiben und viel Geld in die Hand nehmen. PwC-Partner Westermann verweist unter anderem auf das Labelling: «Vereinfacht gesagt, müssen Daten angeschrieben sein, damit sie vom Algorithmus überhaupt erkannt werden.» Das gilt auch, um Machine Learning voran-



zubringen, von dem man sich heute Verbesserungen im Umgang mit Big Data verspricht. Ein simples Beispiel, das in dem Zusammenhang oft genannt wird: Will man, dass der Algorithmus Katzen auf Bildern erkennt, muss man ihn mit Hunderten von Katzenbildern füttern und die Bilder mit der Information versehen, dass darauf eine Katze zu sehen ist. So lernt der Algorithmus das Tier von anderen zu unterscheiden. Ein sauberes Labelling brauchen vor allem Daten ohne klare Struktur. Neben Bildern können das auch Texte in PDF-Dateien oder Sequenzen von Überwachungskameras sein.

Die Grossbanken kennen das Problem. Westermann: «Wenn ich in Zukunft besonders auffällige Transaktionen erkennen möchte, muss ich der Maschine erst aus meinen historischen Transaktionen zeigen, auf welche das zutrifft.» Das versuchte auch Signac, ein Joint Venture von Credit

Suisse und Palantir. Die Grossbank und der Softwarekonzern, der auch mit der CIA zusammenarbeitet, überwachten auf diese Weise die CS-Angestellten, um «allfälliges internes Fehlverhalten» aufzuspüren. Doch wie die CS jetzt bestätigt, wurde Signac nach nur einem Jahr wieder begraben. Grund dafür seien neue Risikokapitalvorschriften, sagt ein Sprecher, weil Signac unter anderem genutzt worden sei, um Risikokapitalberechnungen zu tätigen. Die Zusammenarbeit habe sonst einwandfrei funktioniert. Mit Facebooks Datenskandal, in den auch Palantir involviert ist, habe die Trennung nichts zu tun.

Genauso wie Google und Facebook hat Palantir einen überragenden Vorteil gegenüber vielen älteren Firmen. Der Konzern ist in der Frühphase der Digitalisierung mit dem Zweck entstanden, mit Daten Geld zu verdienen. Das Geschäftsmodell ist glasklar. «Wenn du Big Data nicht zum Kern deines Schaffens machst, wird das Ganze zum Innovationstheater», sagt Digitalprofi Patrick Comboeuf, der für die SBB, Swiss

Life und Ifolor tätig war. Er zitiert dazu gerne die US-Informatikerlegende Grady Booch: «A fool with a tool is still a fool.»

Eine Erkenntnis, zu der auch Paul Hawtin gekommen ist. Der Gründer des «Twitter Hedge Funds» Derwent Capital Markets warb Kunden mit dem Versprechen an, den Kurs des Dow Jones mit Twitter-Daten vorauszusagen. Aus relevanten Tweets wollte er Stimmungen ►



► ausmachen, die das Anlegerverhalten um drei Tage vorwegnehmen sollten. Der Fonds überlebte nur einen Monat. Gegenüber der «Financial Times» versicherte Hawtin, dass seine Big-Data-Maschine derart gut laufe, dass er daraus ein Massenmarktprodukt machen werde. Es entstand eine Online-Trading-Plattform, deren Spur sich inzwischen verloren hat.

Finanzprofessor Karl Schmedders von der Uni Zürich hat mehrere Big-Data-getriebene Fintech-Firmen beobachtet und stellt einen Vergleich zur Jagd auf: Früher habe man die eine Kugel im Gewehr mit viel Bedacht abgefeuert, «heute schießt man mit Schrot in den Baum rein, und irgendwo fällt dann ein Vogel runter».

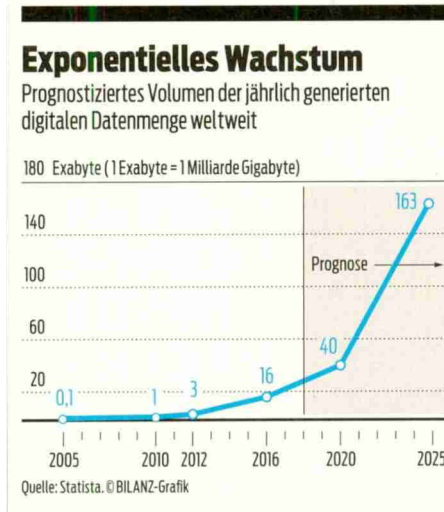
Bei der Migros, die auf einem beispiellosen Datenschatz sitzt, will man genau das verhindern. Dank seines Kundenbindungsprogramms Cumulus gilt der Retailer als interessantes Anschauungsbeispiel in Sachen Big Data. Weil die Migros aber nicht

wie Google eine Internetfirma, sondern ein 90 Jahre alter Detailhändler mit angehängter Produktion ist, hat man einen eher semiprofessionellen Zugang zu den Daten.

Das wurde 2017 an einem Podiumsgespräch der Schweizerischen Akademie der Technischen Wissenschaften deutlich. Dort soll Jan Vonderlinn, Leiter personalisiertes Marketing und Datenschutzbeauftragter, gesagt haben, die Migros habe bislang sechs bis acht Milliarden Franken ins Cumulus-Programm investiert, könne aber erst seit zwei Jahren «Vernünftiges» damit anfangen. Darauf angesprochen, winkt die Migros ab. Es handle sich um grobe Schätzungen, die an einem Podium geäußert und aus dem Kontext gerissen worden seien. Bei einer Preisverleihung Ende Juni liess Ex-Migros-Chef Herbert Bolliger aber durchblicken, dass Vonderlinns Aussagen nicht ganz falsch sein könnten: «Beim Thema Big Data stehen wir noch am An-



fang. Letztlich müssen die wissenschaftlichen Analysen auch praktisch umgesetzt werden. Die Datentöpfe sind voll, müssen aber noch verknüpft werden.»



Da kommen die «Retter» von Big Data ins Spiel: Machine Learning und die künstliche Intelligenz. Wenn die Rechner anfangen, selber zu lernen und zu denken, trennt sich die Spreu vom Weizen, so die Hoffnung. Die neuen Schlagworte sind nun zwar ins Standard-Vokabular jedes Gründers und Managers eingeflossen, doch wirklich damit arbeiten können nur wenige. Teralytics, ein ETH-Spin-off, das mit Handydaten Mobilitätsmodelle erstellt, scheint hier auf vielversprechendem Weg zu sein. «Bei komplizierteren Datensätzen können Machine Learning und künstliche Intelligenz nun sehr viele Fragen beantworten», sagt Co-Gründer Georg Polzer.

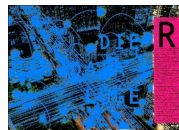
### FALSCHER REALITÄT

Teralytics zeigt Stadtplanern und Verkehrsbetrieben auf, wie sich die Menschen im öffentlichen Verkehr bewegen. Solche Modelle lassen sich sonst nur mit viel Aufwand und Manpower erstellen. Ihre anonymisierten Bewegungsdaten beziehen sie von den Telekom-Unternehmen. Aber auch hier gibts viele Stolpersteine: «Man

muss höllisch aufpassen, dass man den Algorithmus nicht mit schiefen Daten füttert und damit eine falsche Realität aufbaut», erklärt Polzer. «Darum müssen unsere Daten sämtliche sozialen Schichten repräsentieren.»

Die grosse Frage wird sein, ob sich das lohnt. Auch Teralytics schreibt im siebten Jahr rote Zahlen und versucht mit Hilfe von Investorengeldern, Marktanteile zu erobern. Selbst etablierte Tech-Grössen, die ihr Geschäftsmodell auf Big Data aufbauen, sind noch längst nicht über den Berg: Uber schliesst Jahr für Jahr negativ ab, zuletzt mit einem Rekordverlust von 4,5 Milliarden Dollar. Spotify verbrennt in jedem Quartal zwischen 100 und 700 Millionen Dollar.

Und es wird nicht einfacher. Im Mai traten in der EU neue Regeln in Kraft, die den Menschen mehr Kontrolle über ihre Daten geben. Die neue Datenschutzverordnung DSGVO verlangt, dass die Informationen noch besser anonymisiert und verschlüsselt werden. Vieles ist noch nicht ausformuliert, klagende Konsumentengruppen könnten das grosse Sammeln zusätzlich erschweren. Die Qualität der Daten wird dadurch nicht besser. ■



Um Grippewellen zu registrieren, ermittelt man Arztbesuche. Das dauert. Google wollte schneller sein und verarbeitete Grippe-Suchanfragen zu einem Big-Data-Frühwarnsystem. Nach vier Jahren wurde es kommentarlos eingestellt. Man war zwar schneller als die herkömmliche Methode - dafür waren aber die Prognosen komplett daneben.

## Glossar

**Big Data** beschreibt eine sehr hohe Anzahl Daten, die unzusammenhängend und schnelllebig sind und dadurch nicht mit den herkömmlichen Mitteln der Informationsverarbeitung erfasst werden können. Die Verbreitung des Internets und die stark zunehmende Rechnerleistung erhöhen die Chancen, mit Big Data umzugehen.

**Algorithmen** sind Grundbausteine in der Software, die Anweisungen geben, wie Aufgaben in elektronischen Geräten erfüllt werden müssen. Berühmtestes Beispiel: die Google-Suche.

**Data Mining** ist ein Verfahren, um relevantes Wissen aus grossen Datensätzen zu ziehen. Dabei

sortieren und modifizieren Algorithmen die Daten. Allerdings lassen sich damit keine Kausalitäten ermitteln, sondern nur Zusammenhänge, die auch zufälliger Natur sein können.

**Künstliche Intelligenz (KI)** vertritt die Idee, Computer Anspruchsvolles erledigen zu lassen. Im weitesten Sinne ist es der Versuch des Menschen, das Gehirn und seine Funktionen künstlich nachzubilden.

**Machine Learning (ML)** ist ein Teilgebiet der KI und beschreibt Software, die ohne zusätzliche Programmierung zum selbständigen Verarbeiten grosser Datenmengen eingesetzt wird. Algorithmen ermöglichen dem Computer zu lernen, wie es der Mensch auch tut - einfach viel schneller. Grossbanken und Geheimdienste nutzen ML, um

**Auffälligkeiten zu erkennen.**

Data Scientists setzen sich mit der Sammlung und Verarbeitung grosser Datensätze auseinander. In der Data Science verschmelzen Statistik, Datenanalyse und Machine Learning. Der Data Scientist befasst sich mit Algorithmen genauso wie mit wissenschaftlichen Methoden alter Schule.

**Nette Tools für  
Schweizer Grössen**

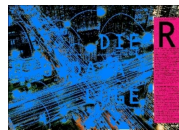
Novartis sieht Daten als «eines der wertvollsten Assets, die wir als Firma haben». Darum ist das Ziel unter CEO Vas Narasimhan: «Becoming a medicines and data science company.» Novartis beschäftigt bereits «Hunderte» Data Scientists. Konkurrentin Roche hat zuletzt die Firmen Flatiron und Foundation Medicine übernommen, die vielversprechende Big-Data-Anwendungen in der Krebsforschung mitbringen.

Machine Learning und künstliche Intelligenz (KI) sind bei der UBS «ein Kernelement unserer Technologiestrategie». So bereitet die UBS etwa einen Chatbot vor, der IT-Probleme lösen soll. Auch werden Tools benutzt, um Stellen intern zu besetzen, mit

den Kunden zu interagieren oder Analysten im Research zu unterstützen. Zudem hat die Bank ein KI-Überwachungssystem zur Früherkennung von Geldwäscherei oder Betrugsfällen im Trading eingerichtet. Mit ähnlichen Tools arbeitet auch die CS.

Die Swisscom nutzt im Kundenservice KI-Anwendungen, um die Angestellten zu entlasten. Die Programme beantworten einfache Anfragen oder unterstützen die Mitarbeiter im Callcenter. Ende 2018 soll ein Chatbot lanciert werden, der sogar selbständig Produktberatungen durchführt. Schon seit zwei Jahren sind Fernbedienungen von TV-Kunden mit einer Sprachsteuerung ausgestattet.

Mit je rund drei Millionen verteilter Kundenkarten haben Migros und Coop Einblick ins Konsumverhalten von drei Vierteln aller Schweizer Haushalte. Ihre Loyalitätsprogramme bergen den wohl wertvollsten Datenschatz der Schweiz. Der Umgang damit ist aber nicht einfach. Was man bislang verarbeiten konnte, wird genutzt, um das Sortiment anzupassen und Angebote zu machen, die besser auf den Kunden zugeschnitten sind.



# „Das Öl braucht noch seinen Motor“

ETH-Professor Dirk Helbing  
stellt beim Umgang mit  
Big Data noch technologische  
und ethische Mängel fest.  
Er plädiert für Kooperation  
statt Konkurrenz.



**Dirk Helbing** (53) ist Professor für Computational Social Science an der ETH Zürich. Der Physiker, Mathematiker und Soziologe mit unverkennbarem Haarschnitt befasst sich unter anderem mit Verkehr, Finanzsystemen, Spieltheorie und Digitalisierung. Er ist beehrter Gast an Podiumsdiskussionen.



**E**s gibt kaum profitable Big-Data-Anwendungen. War alles nur ein grosser Hype?

Das glaube ich nicht. Wir sind jetzt in der Phase, in der wir merken, dass Big Data nicht einfach vollautomatisch funktioniert, sondern dass es da wesentlich mehr Erfahrung und Wissenschaft braucht, als man gedacht hat.

#### Woran scheitern Big-Data-Projekte?

Es braucht genügend wissenschaftliche und ethische Kompetenz an Bord. Und einen interdisziplinären und multiperspektivischen Ansatz. Man hat die Schwierigkeiten unterschätzt, die mit grossen Datenmengen einhergehen. Dazu zählen Scheinkorrelationen, also Muster, die man in den Daten findet, die aber keine Bedeutung haben.

#### Es heisst, mit Machine Learning und künstlicher Intelligenz (KI) bekomme man diese Probleme in den Griff.

Wenn Big Data das neue Öl sein soll, dann haben wir den Motor dazu noch nicht erfunden. Viele hoffen nun, dass die künstliche Intelligenz dieser Motor ist. Die grosse Kunst ist, die Rohdaten zu destillieren und zu veredeln, also in nützliche Informationen und Wissen zu verwandeln. Allerdings gibt es auch hier einige unerwartete Hindernisse. Denn man hat festgestellt, dass KI nicht zwangsläufig objektiv ist. Sie lässt sich manipulieren und kann diskriminieren. Die Fehler, die man dem Menschen anlastet, verschwinden nicht einfach mit der künstlichen Intelligenz.

#### Wie kann man Big Data manipulieren?

Viele dieser Algorithmen reagieren auf Echtzeitdaten. Sie fliessen ständig ins

System ein und werden zum Lernen verwendet. Wenn sie manipuliert sind oder falsch bearbeitet werden, entstehen Fehler. Im Moment schaffen es lediglich Google, Facebook oder Amazon, mit riesigen Datensätzen via Werbeinserate Geld zu verdienen.

#### Haben andere und kleinere Unternehmen überhaupt eine Chance?

Die Giganten haben alle einige Jahre Minusgeschäfte gemacht. Das ist bei europäischen Unternehmen vielleicht nicht so üblich. Man braucht einen langen Atem. Ausserdem gab es damals, als diese Geschäftsmodelle entstanden, noch nicht so viel Konkurrenz. Heute stehen wir mit diesen Konzernen im Wettbewerb. Europa muss sich gegenüber den USA und China differenzieren, also eigene Marktnischen finden. Ich empfehle eine Vernetzung, also im Sinne von kombinatorischer Innovation einen gemeinsamen Datenpool für Europa zu schaffen. Die gesammelten Informationen über uns würden an Daten-

postfächer gesendet, die wir selber verwalten. So liesse sich auch die informationelle Selbstbestimmung umsetzen.

#### Was heisst das?

Ich spreche von digitalen Assistenten, die uns helfen, unsere Daten zu verwalten, sodass wir entscheiden können, wer welche Daten wofür verwenden darf. Das hätte Vorteile für die Zivilgesellschaft und die Demokratie, denn wir bekämen die Kontrolle über unsere Daten zurück. Die Wirtschaft hätte den Vorteil, dass alle Firmen, die das Vertrauen gewinnen können, Zugang zu den

Daten hätten - also auch Spin-offs und KMUs.



*Das Problem ist nur, dass die Konzerne,  
die am meisten Daten über uns haben,  
diese nicht herausrücken wollen.*

*Da hapert es mit der Transparenz.*

Das verlangsamt die Erschliessung und die Veredelung der Daten. Angesichts der raschen Entwicklung kann man sich solche Verzögerungen nicht erlauben. In der Regel gewinnt der Schnellste. Eine rein kompetitive Strategie, wie sie aus den Zeiten der alten, materiellen Wirtschaft stammt, in der die Ressourcen begrenzt sind, ist nicht zeitgemäss, nicht zielführend. Die Digitalisierung erfordert ein Umdenken. Kooperation ist Trumpf.