BIG DATA

Wie wir im Windschatten der Geheimdienste ausgespäht und manipuliert werden und andere Aspekte über Big Data

von Werner Mittelstaedt

Vorbemerkungen

Im Windschatten des NSA-Skandals werden die vielfältigen „ganz legalen“ Ausspähungen sowie Manipulationen vieler Millionen Menschen durch Algorithmen, die „Big Data“ nutzen, von großen Teilen der Bevölkerung nicht richtig wahrgenommen und/oder nicht ernst genug genommen. Die Politik agierte und reagierte bislang zu zaghaft und hinkt den vielfältigen Entwicklungen um „Big Data“ ohne Zweifel hinterher. Sie kann zwar diese Entwicklung mit großer Wahrscheinlichkeit nicht mehr aufhalten, aber trotzdem kann sie mehr für die Sicherheit unserer Daten unternehmen. Doch letztendlich liegt es an uns, die „Hoheit“ über unsere Daten zu bewahren.

Nachfolgend werden Gegenwart und Zukunft von „Big Data“ grob skizziert. Es wird hinterfragt, ob wir in Zukunft den Ausspähungen und Manipulationen durch kleine und mittlere Unternehmen bis hin zu global agierenden Konzernen, aber auch denen von Behörden und Staaten entkommen können, die das riesige Informationspotential von „Big Data“ für ihre ökonomischen und manipulativen Zwecke oder zur Kontrolle über Menschen nutzen. Durch Algorithmen, die die Daten des „Big Data“ nutzen, werden unsere Grundrechte verletzt und der Diskriminierung einzelner Personen und bestimmter Gruppen wird Tür und Tor geöffnet. Andererseits liefern die Algorithmen, die „Big Data“ nutzen, viele interessante Möglichkeiten die Zukunft zu gestalten.

Big Data im Kontext der Ausspähaffären

Spätestens durch die qualitativ und quantitativ ungeheuerlichen Enthüllungen durch Edward J. Snowden wissen wir, dass wir auf vielfältige Art und Weise durch den NSA und andere Geheimdienste ausspioniert und unter Generalverdacht für kriminelle und für terroristische Aktivitäten unterschiedlichster Provenienz gestellt werden. Noch immer werden wir über neue Formen der Überwachung durch die große Anzahl von Dokumenten in Kenntnis gesetzt, die Edward J. Snowden der Presse und damit der Weltöffentlichkeit lieferte, obwohl seine Enthüllungen bald ein Jahr alt sind. Die Beschwichtigungsrhetorik der US-Administration lässt kaum hoffen, dass die Ausspä-

hungen von ungezählten Menschen, Firmen und Konzernen, die in vielerlei Hinsicht ungesetzlich, undemokratisch und menschenrechtsverletzend sind, durch den NSA und anderer Geheimdienste eingestellt werden. Darüber regt sich, völlig zu Recht, ein großer Teil der Menschen auf. Unverständlicherweise ist seit einigen Wochen auch die Bundesregierung nicht mehr an ein No-Spy-Abkommen mit den USA interessiert. Dies verwundert sehr, weil die Entrüstung, insbesondere über das abgehörte Kanzlerinnen-Handy, zuerst sehr groß war. Stattdessen strebt die Bundesregierung nun ein Cyberdialog mit den USA an. Das bedeutet letztendlich, dass sich an den Ausspähungen durch den NSA auch in Zukunft substantiell nichts ändern wird.

Viel weniger oder fast keine Aufregung gibt es „noch“ über die technologischen Möglichkeiten, die vorhanden sind und bereits vielfältig genutzt werden, um Menschen(massen) auszuspähen und zu manipulieren. Letzteres geschieht überwiegend „ganz legal“ mit Daten, die wir als Nutzer des Internets, von Smartphones und Handys, durch das Führen von Telefongesprächen, durch bargeldloses Bezahlen von Waren und Dienstleistungen mit Kreditkarten, als Mitglied von sozialen Netzwerken im Internet oder auch durch die schlichte Benutzung von Suchmaschinen im Internet hinterlassen. Zudem werden wir durch Techniken wie Videoüberwachung, Bilderkennungssoftware und anderen biometrischen Identifikationsmöglichkeiten sowie dem Internet der Dinge, z. B. durch die automatische Standortidentifikation mittels RFID-Systeme (engl. radio-frequency identification) und durch das Verbinden zahlreicher Gegenstände wie Kühlschränke, Wohnungs- und Hausüberwachungsanlagen, andere Elektronik in Wohnungen und Häusern, „Black Boxen“ in Autos, die mit dem Internet verbunden sind, potentiell ausgespäht und/oder überwacht. Eine Gesichtserkennungssoftware ist in der Datenbrille „Google Glass“ enthalten, die aber noch nicht verkauft werden darf, weil ernsthafte Datenschutzbedenken einen Verkauf verbieten. Insgesamt entsteht hier erst ein großer Markt, um Dinge aus unserer Arbeits- und Lebenswelt „smart“ zusammenzuführen und sie mit dem Internet zu verbinden. Dafür haben sich die Begriffe wie Smart Home, Smart Glass, Smart City und Smart Grid bereits etabliert.

Es stehen vielfältige technologische Möglichkeiten zur Verfügung, um das riesige Informationspotential dieser gewaltigen Datenmengen, die „Big Data“ genannt werden, zu nutzen. Die riesigen Daten von „Big Data“ werden schon längere Zeit für ökonomische und politische Ziele, aber auch für wissenschaftliche Erkenntnisse genutzt. Sie bieten interessierten Kreisen „ganz legale“ Chancen, um Informationen über Menschen (Habitus, Konsum- und Freizeitverhalten, Freundes- und Bekanntenkreis u.v.a.) zu gewinnen. Durch die fortgeschrittenen Realisierungsmöglichkeiten durch Software mit immer besseren Algorithmen im Verbund mit preisgünstigen und sehr leistungsfähigen Computern und billigem Speicherplatz hat sich um „Big Data“ in den letzten Jahren eine boomende Branche entwickeln können.

Was genau ist „Big Data“?

Es sind im Grunde genommen sämtliche Daten, die elektronisch erfasst und gespeichert werden. Insbesondere sind es die gewaltigen Datenmengen des World Wide Web (Internet) inklusive seiner Logdateien sowie die Daten auf Handys und Smartphones, die mit Servern in der ganzen Welt verbunden sind. Darüber hinaus die Protokolle von Telekommunikationsverbindungen und außerdem alle anderen Geräte, die Daten erfassen und speichern können und sich, ähnlich wie die Daten des Internets, durchsuchen und analysieren lassen. Es können auch Daten und Datenbanken sein, die sich auf einem einzelnen Server befinden und nicht öffentlich zugänglich sind. Sie können schon lange mit dem riesigen Datenbestand von „Big Data“ des Internets und ggfs. weiterer Datenquellen auf Korrelationen abgeglichen und für diverse Anwendungen Datenmaterial liefern. (Eine Korrelation beschreibt eine Wechselbeziehung zwischen mehreren Merkmalen, Ereignissen, Zuständen oder Funktionen. Zwischen ihnen muss nicht zwingend eine kausale Beziehung bestehen. Es können auch zufällige Beziehungen zwischen den Elementen bestehen.)

Das Volumen der Datenmenge von „Big Data“ ist so groß, dass wir es verstandesmäßig nicht mehr erfassen können, denn es handelt sich um sehr viele Exabyte. Ein Exabyte hat eine Trillion (10^{18}) Bytes oder eine Milliarde Gigabyte oder eine Million Terabyte oder Tausend Petabyte. Für das Jahr 2013 wurde die Gesamtmenge von „Big Data“ auf 1.200 Exabyte geschätzt (Mayer-Schönberger und Cukier 2013, S. 16). Andere Schätzungen gehen noch weit über diesen Wert hinaus. Um die hier genannte Datenmenge einigermaßen mit dem Verstand zu erfassen, folgendes Zahlenspiel: Mein Buch „SMALL – Warum weniger besser ist und was wir dazu wissen sollten“ aus dem Jahre 2012 hat ein Umfang von 266 Seiten. Das Buch hat als PDF-Format eine Datenmenge von ca. 3.000 Kilobyte (3.000.000 Byte). Das bedeutet, dass *jeder* der 7 Milliarden Menschen auf der Erde ca. 57000 (!) Bücher dieses Umfangs besitzen müsste, um die Datenmenge von „Big Data“ des Jahres 2013 zu erreichen. Egal, ob die geschätzte Datenmenge von „Big Data“ des Jahres 2013 einigermaßen richtig abgeschätzt wurde oder kleiner war. Fest steht, dass es sich um eine unfassbare große Menge an Daten handelt.

Schätzungen gehen davon aus, dass sich das globale Volumen von „Big Data“ etwa alle zwei bis drei Jahre verdoppeln wird. Dann müsste spätestens im Jahr 2016 jeder Mensch auf der Welt die doppelte Anzahl der oben genannten Bücher besitzen. Das muss nicht verwundern, weil ununterbrochen und mit steigernder Geschwindigkeit Zeitschriften, Tageszeitungen, Enzyklopädien, der überwiegende Teil wissenschaftlicher Arbeiten, unzählige Bücher, die Einträge ins Facebook und anderer sozialer Netzwerke im Internet, Behördenprotokolle, riesige Datenbanken aus Wirtschaft, Wissenschaft, Forschung und Politik, Webseiten privater Personen bis hin zu Konzernen im Internet hinzugefügt werden.

Einerseits kann und wird „Big Data“ für ökonomische Interessen und für politische Ziele dahingehend eingesetzt, um Menschen, vergleichbar mit den Methoden der NSA und anderer Geheimdienste, auszuspähen, ihr Konsumverhalten zu manipulieren und damit letztlich auch die Privatsphäre einzelner Menschen zu verletzen.

Andererseits können die Algorithmen, die Softwareentwickler erstellen, um aus „Big Data“ Informationen zu generieren, für den Einzelnen und Gesellschaften außerordentlich nützlich sein. Weil „Big Data“ dazu geeignet ist, wertvolles Wissen für ökonomische, gesellschaftliche, soziologische und ökologische Zwecke zu generieren, muss die technologische Nutzung von „Big Data“ als hochgradig ambivalent eingestuft werden. Vieles ist gut an der Nutzung von „Big Data“, vieles aber schlecht und kann gefährlich werden. Letzteres deshalb, weil für einige Nutzer der algorithmischen Auswertungen von „Big Data“ Korrelationen als wertvolleres Wissen interpretiert werden können, als wirkliche Fakten oder kausale Erklärungen, die sich in der Zukunft falsifizieren lassen. So können zum Beispiel über Algorithmen aus „Big Data“ Informationen über Personen gewonnen werden, um vermeintlich ihre Bonität als Kreditkunden sicher zu stellen. Durch die Algorithmen können aber letztlich keine realen Fakten über die Bonität von Kreditkunden gewonnen werden, sondern nur Annahmen oder falls Personenprofile erstellt werden, möglicherweise nur „Sympathiewerte“ aufgrund der ermittelten „Fakten“ aus den Korrelationen. Ein „realer Fakt“ zur Absicherung der Bonität eines Kunden wäre z. B. ein regelmäßiges Einkommen oder bislang abgezahlte oder nicht abgezahlte Kredite. Im Jahr 2013 gab es bei der Wirtschaftsauskunftei Schufa Überlegungen, ob Facebook-Daten für die Bewertung der Bonität von Kreditkunden genutzt werden sollten. Diese Überlegungen wurden wegen der Einwände von Datenschützer und Politiker zunächst fallen gelassen.

Sicher ist, dass sich über Algorithmen, die „Big Data“ als Datenquelle nutzen, Lebenszusammenhänge bis hin zu konsistenten Profilen (digitale Profile) einzelner Personen bereits anfertigen und sich Interessengruppen mit den dazugehörigen Personenkreisen identifizieren lassen. Untersuchungen haben gezeigt, dass viele digitale Profile das Denken und Handeln sowie das persönliche Umfeld von Menschen bis ins Detail richtig wiedergeben werden – insbesondere dann, wenn eine Person viel bis sehr viel über sich im Internet an Daten hinterlässt.

Der Unterschied bei der nicht geheimdienstlichen Nutzung von „Big Data“ durch Algorithmen gegenüber den Ausspähmethoden der Geheimdienste ist die Tatsache, dass keine E-Mails, keine SMS, MMS gehackt werden und keine Telefongespräche ab- oder -mitgehört werden.

„Big Data“ ermöglicht uns das „Weltwissen“ und „aktuelle Informationen über alles Mögliche“ das im Internet gespeichert ist, über verschiedene Systeme, z. B. PC, Notebook, Smart-Phone, Tablet PC zu nutzen. Wir können mit sozialen Netzwerken wie Facebook oder LinkedIn kommunizieren, über das Internet Produkte und Dienstleistungen erwerben bzw. in Anspruch nehmen, was noch vor nicht allzu langer Zeit nur in realen Geschäften mit realen Firmen möglich war, die nun über die virtuellen Welten des Internets ihre Geschäfte tätigen. Wir können Preisvergleichsmaschinen nutzen und so weiter. Jede Preisvergleichsmaschine im Internet wertet über Algorithmen das Internet aus und zeigt die gesuchten Produkte oder Dienstleistungen für den Nutzer an. Ähnlich wird das „Big Data“ des Internets für Dienstleistungen wie die Suche nach Gebrauchtwaren, Hotels u.v.a. mittels Algorithmen durchsucht und ausgewertet. Wir können natürlich „Big Data“ über Suchmaschinen nach unseren eigenen Vorstellungen durchsuchen.

Firmen und Konzerne beauftragen Softwareentwickler, die Algorithmen entwickeln, um „Big Data“ z. B. nach Kritiken über Produkte oder Dienstleistungen mit dem Ziel abzusuchen, um Stärken oder Schwächen eines Produktes über Meinungen und Bewertungen von Verbrauchern zu identifizieren. Diese Auswertungen sollen dann Produkte und Dienstleistungen verbessern. Darüber hinaus können mit „Big Data“-Auswertungen Marketingstrategien optimiert werden.

Unternehmen werten mit Software „Big Data“ aus, um sich im Wettbewerb mit Konkurrenten zu behaupten und auch zusätzliche Informationen über sie zu gewinnen, die sonst so nicht zu haben sind. Ebenso wird durch „Big Data“ Marktforschung betrieben, etwa um herauszufinden, welche Eigenschaften ein spezifisch neues Produkt für ein bestimmtes Unternehmen haben sollte und ob es Chancen hat, sich auf dem Markt durchzusetzen. Natürlich können die Aktivitäten von Konkurrenten über „Big Data“ ergänzend zu herkömmlichen Informationsquellen ausgewertet werden.

Nachfolgernd werden weitere Nutzungen von „Big-Data“ grob skizziert:

- Auswertungen von „Big Data“ über den Kauf bestimmter Medikamente, z. B. gegen Grippe, um herauszufinden, ob beispielsweise eine Grippewelle droht. Viktor Mayer-Schönberger und Kenneth Cukier schrieben dazu: „ […] Google verglich die 50 Millionen am häufigsten von US-Bürgern eingegebenen Suchbegriffe mit den Daten der CDC [Anm. W.M.: Center for Disease Control and Prevention] zur Ausbreitung der jährlichen Grippeepidemien von 2003 bis 2008, um so eine Korrelation zwischen Suchanfragen und befallenen Gebieten zu ermitteln. Das hatten auch andere schon versucht, aber niemand verfügte über so viele Daten, so viel Rechnerleistung und so großes statistisches Können wie Google. […] Das System suchte stattdessen nur nach Korrelationen zwischen der Häufigkeit bestimmter Suchbegriffe und der Ausbreitung der Grippewelle über Zeit und Raum. Insgesamt wurde die enorme Zahl von 450 Millionen unterschiedlicher mathematischer Modelle auf ihre Tauglichkeit geprüft, wobei jeweils die Voraussagen mit den tatsächlichen Grippedaten der CDC von 2007 und 2008 verglichen wurden. Und so fanden die Google-Entwickler tatsächlich das richtige Modell, das bei 45 Suchbegriffen eine starke Korrelation zwischen der darauf basierenden Grippevorhersage und den amtlichen landesweiten Zahlen zur Verbreitung der Epidemie aufwies. Google konnte damit die Ausbreitung der Grippe genauso gut wie die CDC feststellen, aber nicht mit ein oder zwei Wochen Verspätung, sondern praktisch unmittelbar. Während der H1N1-Krise des Jahres 2009 erwies sich das Google-System daher als nützlicherer und schnellerer Indikator als die Regierungsstatistiken mit ihren unvermeidlichen Verzögerungen. […] Neu ist dabei, dass Google keine Gewebeproben einsammelt oder Berichte von Hausärzten auswertet. Stattdessen beruht die Methode auf ‚Big Data‘ – der Fähigkeit, Informationen so zu nutzen, dass neue Erkenntnisse, Güter oder Dienstleistungen von bedeutendem Wert gewonnen werden. Mit dieser Methode verfügt die Menschheit über ein neues Instrument, um im Falle einer Pandemie die Ausbreitung

vorauszusagen und damit zu verhindern.“ (Viktor Mayer-Schönberger und Kenneth Cukier 2013, S. 8-9).

- Über „Big Data“ wird das Kaufverhalten manipuliert. Christian Buck berichtete darüber in der Technology Review: „David Petersen weiß ziemlich genau, wo sich seine amerikanischen Mitbürger gerade aufhalten und welche persönlichen Vorlieben sie haben – zumindest jene 90 Millionen, deren Smartphones die Server von Sense Networks regelmäßig mit Informationen über ihren Aufenthaltsort versorgen. Das von ihm geleitete Unternehmen mit Sitz in New York hat sich darauf spezialisiert, aus den Bewegungsmustern von Menschen Schlüsse auf deren Vorlieben und Lebensgewohnheiten zu ziehen. Solche Erkenntnisse bringen die Werbebranche ihrem großen Traum wieder ein Stück näher: Potenziellen Käufern die richtige Werbung zum richtigen Zeitpunkt auf dem Handy zu servieren und damit einen todsicheren Kaufanreiz zu schaffen. Sense Networks nutzt dafür die Fähigkeit moderner Smartphones, mithilfe ihrer GPS-Empfänger jederzeit die Position des Nutzers auf wenige Meter genau bestimmen zu können. Werbefinanzierte Apps auf dem Mobiltelefon senden diese Ortsinformationen immer dann an die Server des Unternehmens, wenn sie von dort neue Werbebanner abrufen – zusammen mit einer Gerätekennung wie Apples IDFA (Identifier for Advertisers), mit deren Hilfe sich jedes Gerät eindeutig wiedererkennen lässt. Mit anderen Worten: Sense Networks weiß nicht nur, wo sich ein Handy im Moment aufhält, sondern kann über einen längeren Zeitraum auch Verhaltensmuster seines Besitzers erkennen. Das Programm registriert, wer regelmäßig beim Handelsriesen Wal-Mart einkauft. Oder es speichert, wann der Besitzer eines Smartphones in einem Stadion sitzt, und verknüpft die Information mit einem Event, das dort stattfindet. Läuft gerade ein Lady-Gaga-Konzert, dürfte es sich um einen Fan der Sängerin handeln. Und wer ständig Inlandsflüge bucht, ist sehr wahrscheinlich ein Geschäftsreisender. ‚Wir haben heute Tausende solcher Merkmale, mit denen wir Smartphone-Nutzer charakterisieren können‘, erklärt Petersen. Mit der Zeit entsteht über den Besitzer des Mobiltelefons ein immer detaillierteres Bild – und das ist Gold wert: Unternehmen können jetzt gezielte Kampagnen fahren, die sich an spezifizierte Zielgruppen richten – etwa an Menschen, die Hamburger lieben und gerade neben einer McDonald’s-Filiale stehen. Mehr noch: Die Algorithmen von Sense Networks erkennen sogar persönliche Zeitmuster potenzieller Kunden. Wer statt samstags lieber schon am Freitag einkauft, kann pünktlich vor seiner Shopping-Tour mit Handy-Werbung beglückt werden.“ (Quelle: www.heise.de/tr/artikel/Big-Data-Is-Watching-You-1808252.html vom 27.02.2013).

- Beeinflussung von Kunden im E-Commerce via Upselling. Das ist die Bezeichnung im Vertrieb für das Bestreben eines Anbieters, dem Kunden statt einer günstigen Variante im nächsten Schritt ein höherwertiges Produkt oder eine höherwertige Dienstleistung anzubieten.

- Beeinflussung von Kunden im E-Commerce durch Produktempfehlungen, die bei der Bestellung empfohlen werden und im Kontext der bestellten Waren stehen. Sie werden durch das Kundenkonto und Kaufverhalten des Kunden vorselektiert.

- Durch „Smart Metering“ werden Energie- und Wasserverbräuche kostengünstig ausgelesen und abgerechnet. Zugleich können dadurch auch Energie- und Wasserverbrauchsprognosen erstellt werden.

- Durch „Big Data“ werden neue Möglichkeiten eröffnet, kriminelle Handlungen (z. B. bei Finanztransaktionen) aufzudecken.

- Die medizinische Diagnostik kann von „Big Data“-Auswertungen neue Erkenntnisse gewinnen.

- Gezielte Online-Werbemaßnahmen können durch „Big Data“ effizienter und in sehr kurzer Zeit realisiert werden, weil z. B. Kundenwünsche über „Big Data“ ermittelt werden.
- Natürlich nutzen die Geheimdienste auch algorithmische Auswertungen von „Big Data“ im großen Stil, um z. B. Bewegungsprofile zu erstellen und das Internet auf verdächtige Gruppen oder Einzelpersonen abzusuchen, die terroristische Aktivitäten planen.

- Am GESIS Leibniz-Institut für Sozialwissenschaften in Mannheim wird u. a. „Computational Social Science“ betrieben. Das Institut schreibt darüber: „Die Abteilung Computational Social Science (CSS) wurde im Oktober 2013 unter der Leitung von Prof. Dr. Strohmaier konstituiert. Sie fokussiert sich zum einen auf die Erforschung von neuen Methoden und Algorithmen für die Analyse sozialwissenschaftlicher Phänomene auf Basis von Daten im World Wide Web wie z.B. Daten aus sozialen Medien. Zum anderen bilden die neu erforschten Methoden und weitere sozialwissenschaftliche Erkenntnisse die Basis für die Entwicklung und den Aufbau forschungsbasierter Angebote für die Sozialwissenschaften. Zur Erforschung von Methoden zur Analyse sozialwissenschaftlicher Phänomene anhand von Daten im WWW werden in der Abteilung Algorithmen und neue, nicht-reaktive Methoden für die Sozialwissenschaften entwickelt. Grundlage für diese Arbeiten sind Ansätze aus den Bereichen Machine Learning, Data Mining und Netzwerkanalyse. Aktuelle Schwerpunkte sind die quantitative Analyse politischer Wahlprozesse und -dynamiken anhand von sozialen Mediendaten sowie die quantitative Analyse von sozialwissenschaftlich-relevanten Prozessen aus sozialen Medien oder Logdaten. Der Community- und Nutzungsorientierung wird bei der Entwicklung von existierenden und dem Aufbau von neuen Angeboten für die Sozialwissenschaften eine hohe Priorität eingeräumt. Durch eine Kombination von reaktiven, nicht-

reaktiven, experimentellen und anderen Verfahren gewinnt die Abteilung Erkenntnisse über Anforderungs- und Incentive-Strukturen relevanter Communities. Diese Erkenntnisse werden anschließend in die Weiterentwicklung von bestehenden, aber auch in die Entwicklung von neuen GESIS-Online-Dienstleistungen einfließen. Forschungsbasierte Nutzungsanalysen sorgen für eine effektive Bereitstellung von in der Community benötigten Ressourcen und neue Methoden erhöhen und optimieren die Nutzerpartizipation in den GESIS-Onlineangeboten." (Quelle: http://www.gesis.org/das-institut/wissenschaftliche-abteilungen/computational-social-science März 2014). Der Redakteur von ZEIT-Wissen, Max Rauner, schrieb über diesen neuen Forschungszweig: „ […] Die Forschungsarbeit steht für einen neuen Trend. Die einen nennen ihn Computational Social Science, die anderen Sozialphysik oder Netzwerkforschung. Wissenschaftler aller Disziplinen mischen dabei mit, ihr Ziel ist ambitioniert: Sie wollen soziale Phänomene mit naturwissenschaftlichen Methoden beschreiben. Ob Panikverkäufe an der Börse, die Verbreitung von Gerüchten und Krankheiten oder Glücksgefühle in einer Gruppe – die Wissenschaftler suchen nach einfachen Gesetzmäßigkeiten im sozialen Gewusel. Wie Menschen reisen, sei auf den ersten Blick eine komplizierte Angelegenheit, sagt Dirk Brockmann. Die meisten legen kurze Wege zur Arbeit oder zur Schule zurück, andere fahren auf Dienstreise oder in den Urlaub. Die Forscher hatten mit einem statistischen Durcheinander gerechnet. Doch dann stellten sie fest, dass dieses Verhalten einer einfachen Regel folgte: Die Anzahl derjenigen, die eine bestimmte Strecke am Tag reisen, nimmt mit zunehmender Reisestrecke ähnlich ab wie die Schwerkraft, wenn man sich von der Erde entfernt – Potenzgesetz, Mathematik achte Klasse. ›Das hat uns alle überrascht‹, sagt Brockmann. ›In der Physik sind das die einfachsten Gesetzmäßigkeiten, die man kennt.‹ Die Simulation der Seuchenausbreitung wurde dadurch enorm vereinfacht. Brockmann ist heute Professor in Illinois und ein gefragter Experte für die Vorhersage der Schweinegrippe in den USA. Als Nächstes will er die Bewegungsmuster in Deutschland mithilfe von Geocaching erforschen, eine Art Schatzsuche mit GPS-Geräten. […]" (Quelle: http://www.zeit.de/zeit-wissen/2010/01/Soziale-Netzwerke März 2014).

- Mit dem von vier Studenten der Northwestern University in Illinois entwickelten Programm „Stats Monkey" werden aus „Big Data" Spielberichte über Baseballspiele automatisch erstellt mit entsprechenden Hintergrundanalysen und Interviews. Mercedes Bunz schreibt dazu: „ […] Stats Monkey vereint dabei zwei digitale Techniken: Im ersten Schritt eignet sich der Algorithmus im Netz veröffentlichte Spielstände an; im zweiten ermittelt er aus diesen Spielständen durch einen sogenannten algorithmischen ‚Entscheidungsbaum' die wichtigsten Akteure und den Spielverlauf. Das Ergebnis fügt er dann mithilfe vorgeschriebener Bausteine zu einem Textfragment zusammen: ‚Team X ging früh in Führung und war nicht mehr einzuholen' oder ‚Team Y versuchte, sich wieder zu fangen, aber vergebens'. Auf diese Weise entsteht mit einem Klick ein

> knochentrockener, aber informativer Sportbericht – schneller als ein Mensch je einen einzigen Satz schreiben könnte. […] Zwar lässt sich über die Qualität des nüchternen Textes streiten, nicht jedoch über die Tatsache, dass hier eine Kulturtechnik automatisiert wird, über die wir Menschen bislang exklusiv verfügten: das Verfassen eines Textes und damit auch das Erzählen einer Geschichte. […]“ (Bunz 2012, S. 14-15).

Dies war eine relativ kleine Auswahl von „Big-Data-Anwendungen“. Sie liefern uns Chancen, Produkte zu verbessern; können Kostensenkungen herbeiführen; ermöglichen Fortschritte im Gesundheitswesen; generieren neue Strategien, um Kriminalität zu bekämpfen; liefern Nutzern des Internets zeitnah aktuelle Informationen, etwa um Preisvergleiche aller Art sekundenschnell zu erhalten; geben den Gesellschafts- und Naturwissenschaften neue Informationsmöglichkeiten, um „Antworten“ auf ganz spezifische Fragen zu erhalten und vieles Andere.

Das Datenpotential von „Big Data“ ist dabei, den Rohstoff für eine neue Wachstumsbranche zu liefern, denn die Analysetechnologien und Algorithmen gewinnen Tag für Tag an Bedeutung und dementsprechend haben Softwareentwickler immer mehr damit zu tun. Der Grund: Die riesige Datenmenge von „Big Data“ enthält Informationen, die im harten Wettbewerb des 21. Jahrhunderts für viele Branchen, aber auch in Wissenschaft und Technik, großen Wert haben können. Modernste Softwareentwicklungen, preiswerte Hochleistungsrechner und relativ billiger Speicher ermöglichen dies.

Wo werden die algorithmischen Auswertungen von Big Data gefährlich?

„Big-Data-Algorithmen“ können, wie schon eingangs betont, unsere Privatsphäre und Grundrechte gefährden. Warum? Weil das zentrale Prinzip von Datenschutzgesetzen, dass personenbezogene Daten eines Menschen nicht ohne seine Zustimmung von anderen Personen oder Institutionen verwendet werden dürfen, im Zeitalter von „Big Data“ teilweise oder ganz ausgehebelt wird. Deshalb können und werden über Algorithmen Informationen über Personen (digitale Profile) aufbereitet, die zum Beispiel Versicherungskonzerne, Banken, Arbeitgeber, Behörden oder Universitäten interessieren. Diese digitalen Profile können aus „Big Data“ über persönliche Websites, soziale Netzwerke, Twitter, aus Blogs und anderen „Datensammelstellen“ des Internets, die Menschen nutzen und in denen sie Fakten über sich und/oder Meinungen über irgendetwas eintragen, über algorithmische Auswertungen gewonnen werden. Die Geheimdienste gehen dreister vor, denn sie nutzen zusätzliche als Informationsquellen gehackte E-Mails und hören Telefongespräche ab. Deshalb können die Daten, die wir über die genannten Technologien in „Big Data“ hinterlassen, über berufliche Karrieren, Versicherungsprämien, Bankkredite u.v.a. entscheiden. Sie entscheiden letztlich auch über „den guten Ruf“ eines Menschen und können ihn nachhaltig schädigen.

Darüber hinaus werden wir, wie oben ausgeführt, aus ökonomischen Interessen über „Big Data“-Technologien dahingehend manipuliert, mehr zu konsumieren.

Wie können wir uns schützen?

1. Schon in frühen Schulklassen sollte der sorgfältige Umgang mit den Möglichkeiten, die das Internet und die technischen Geräte bieten, die letztlich „Big Data“ erzeugen (PC, Tablet PC, Smartphone u.a.) erlernt und ein Pflichtfach werden.

2. Allgemein muss jeder, der in „Big Data“ Datenspuren hinterlässt, wenn er etwa Texte in Facebook, Twitter oder irgendwelchen Blogs absetzt, darauf vorbereitet sein, dass diese Texte gegen ihn verwendet werden können. Das gilt auch für das Internet der Dinge, das letztlich Daten im „Big Data“ erzeugt. Drei einfache Beispiele: 1.) Eine „Black Box“ im Auto, die auch den aktuellen Standort z. B. via GPS itracker des Fahrzeugs ermittelt, verhindert potentiell die Anonymität einer Autofahrt. Ganz allgemein sind Autohersteller dabei, das Auto vollständig mit dem Internet zu vernetzen. Das hat z. B. Vorteile in der schnelleren Unfallversorgung, in der das Auto rasch geortet und gefunden werden kann, um erste Hilfe zu leisten. Es gibt auch Überlegungen, das Fahrverhalten des Besitzers durch Telematische Dienste, die mit dem Internet verbunden sind, zu überprüfen. Dadurch ließen sich z. B. Prämien für die Kfz-Versicherung reduzieren. Eine deutsche Versicherung bietet einen entsprechenden Tarif an. Dieser basiert darauf, dass Kontrollgeräte Informationen zum Fahrverhalten an den Versicherer übermitteln. Wer sicherer fährt, zahlt weniger, muss sich aber auch darüber Gedanken machen, dass er damit über persönliche Daten seine „Datenhoheit“ verliert. 2.) Eine moderne Sportuhr für Freizeitläufer und Radfahrer ermittelt den Fitnesszustand. Die Daten können in einem Webportal analysiert werden. Manche Sportuhren liefern auch interessante medizinische Daten und die Strecken, die gelaufen oder gefahren wurden. Letztendlich gelangen die Daten ins „Big Data“ und können dort von interessierten Menschen gehackt werden oder, wenn das Webportal nicht ausreichend geschützt ist, über Algorithmen ausgewertet werden. 3.) Es ist möglich, den Kühl- und Gefrierschrankinhalt in seiner Wohnung über Anwendungen für Smartphones zu kontrollieren. Die Protokolle gelangen auch ins Internet. Dort werden sie für den Anwender ausgewertet, sie können aber potentiell auch für „andere Interessenten“ ohne unsere Zustimmung verwendet werden.

3. Thomas F. Dapp von der „Deutsche Bank Research“ hat in einer Studie über Big Data“ folgendes geschrieben, die meine Argumente unterstreicht: „ […] Wenn z.B. Google eine web-basierte Brille anbietet oder an der Entwicklung autonom fahrender Fahrzeuge forscht, dann umfasst dies mehr als nur das

Angebot eines selbstfahrenden Autos und mehr als nur das Angebot einer mit dem Internet verbundenen Brille. Beide Produkte erlauben einen beinahe schrankenlosen Zugriff auf teils intime und persönliche Daten. Ein einzelnes Unternehmen bekommt Einblicke in die täglichen Handlungsweisen der Menschen, deren Aktivitäten, Neigungen und Identitäten. Eine web-basierten Brille offenbart z.B., wohin die Blicke der Menschen zielen, wie lange und wie oft sie ihre Blicke auf Gegenstände, Werbung oder Menschen richten. Damit erhält das Unternehmen die Möglichkeit, die Big-Data-Analyse-Instrumente optimal in Echtzeit einzusetzen, um wertvolle Muster aus den gesammelten Daten zu ziehen, diese gegebenenfalls zu monetarisieren bzw. Menschen individuelle Werbebotschaften zu senden. Die Begehrlichkeiten der großen Internetunternehmen, diese zusätzlichen Informationen für unterschiedliche Zwecke mehrfach zu nutzen, nehmen zu, während die gesellschaftlichen Folgen (bis jetzt) weitgehend unerforscht sind. […]“ (Quelle: www.dbresearch. de/PROD/DBR_INTERNET.DE-PROD/PROD0000000000328652/Big+Data+-+die+ungez%C3%A4hmte+ Macht.pdf vom 13. März 2014).

4. Wir können uns schützen, wenn wir dem „Big Data“ möglichst wenige persönliche Daten liefern. Das gilt insbesondere für Daten, aus denen digitale Profile erstellt werden können. Dies ist notwendig, wenn wir die Konsequenzen, die aus möglichen Anfertigungen von digitalen Profilen resultieren können, nicht wollen.

5. Wir sollten nur noch verschlüsselte E-Mails benutzen, den PC oder das Note Book mit tagesaktuellen Virenscannern schützen, die Passwörter regelmäßig wechseln sowie verschiedene Passwörter für verschiedene Dienste verwenden. Darüber hinaus können wir den Software- und Hardwaremarkt für abhörsichere Handys und Smartphones unterstützen.

6. Insbesondere können wir uns vor Ausspähungen schützen, indem wir Suchmaschinen nutzen, die IP-Adressen nicht speichern. Wenn wir eine „ganz normale“ Suchmaschine nutzen, werden unsere Suchanfragen gespeichert. Mit Tracking-Cookies werden unsere Suchbegriffe, der Zeitpunkt und die Dauer unserer Suchanfragen und die ausgewählten Links protokolliert, um diese Informationen dann in einer Datenbank zu speichern. Viele Suchanfragen liefern zahlreiche persönliche Informationen. Daraus können unsere Interessen, Familienverhältnisse, politische Überzeugungen, ggfs. unser Gesundheitszustand und vieles andere ermittelt werden. Es sind Informationen für Marketingspezialisten, Behördenmitarbeiter, Hacker und Kriminelle, die in den Besitz unserer persönlichen Suchdaten kommen möchten.

7. Unsere Smartphones können wir sicherer benutzen, wenn WLAN, Cookies, Ortung, Bluetooth und Roaming nur selten eingeschalten werden, also nur

dann, wenn sie wirklich unbedingt benötigt werden. Dadurch erschweren wir Ausspähungen.

Wo bleibt die Politik?

Für die „Wissenschaftlichen Dienste“ des Deutschen Bundestags hat Sabine Horvath den Begriff „Big Data“ beschrieben und ihn kritisch hinterfragt. Dieser hochinformative Text zeigt deutlich die Problematik von „Big Data“ und dem Datenschutz auf: „ […] Neben den unbestritten großen Potentialen von Big Data für Wirtschaft, Wissenschaft und Gesellschaft werden in der zunehmend intensiver geführten Debatte über die neuen Möglichkeiten auch **kritische Stimmen** laut. Denn gerade die Nutzung der für Big Data besonders interessanten personenbezogenen Daten kollidiert mit zentralen **europäischen datenschutzrechtlichen Prinzipien**, wie dem Recht auf informationelle Selbstbestimmung, dem Schutz personenbezogener Daten und der Zweckbindung von erhobenen Daten, kodifiziert in der Europäischen Grundrechtecharta und dem Bundesdatenschutzgesetz. Auch eine Pseudonymisierung oder Anonymisierung von Daten ist hier nur von begrenztem Nutzen, weil die für Big Data typische Kombination vieler Datensätze häufig **eine De-Anonymisierung** ermöglicht. Einige Beobachter richten zudem den Blick auf die möglichen Auswirkungen auf unser **wissenschaftliches Weltbild**, in dem die Ergründung und die Wichtigkeit kausaler Zusammenhänge nun zunehmend durch statistische Korrelationen abgelöst werden könnte. Und schließlich bleibt zu fragen, wo in einer Welt, in der Entscheidungen zunehmend von datenverarbeitenden Maschinen dominiert werden, die **menschliche Urteilsfähigkeit** oder auch Intuition ihren Platz finden kann. Denn diese könnte manchmal auch nahelegen, bei bestimmten Entscheidungen eben gerade nicht der Datenlage zu folgen. […].“ (Quelle: http://www.bundestag.de/dokumente/analysen/2013/Big_Data.pdf Stand März 2014).

Die bislang aufgezeigten Defizite mit der Datensicherheit im Umfeld von „Big Data“ kann und muss Politik entschärfen.

In Deutschland existiert das Grundrecht auf informationelle Selbstbestimmung. „Das Recht auf informationelle Selbstbestimmung ist im bundesdeutschen Recht das Recht des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner personenbezogenen Daten zu bestimmen. Es handelt sich dabei nach der Rechtsprechung des Bundesverfassungsgerichts um ein Datenschutz-Grundrecht, das im Grundgesetz für die Bundesrepublik Deutschland nicht ausdrücklich erwähnt wird. Der Vorschlag, ein Datenschutz-Grundrecht in das Grundgesetz einzufügen, fand bisher nicht die erforderliche Mehrheit. Personenbezogene Daten sind jedoch nach Art. 8 der EU-Grundrechtecharta geschützt.“ (Quelle: de.wikipedia.org/wiki/Informatio nelle_Selbstbestimmung Stand März 2014).

Politik kann u.a. als Antwort auf die Ausspähaktionen im Kontext von „Big Data“ folgendes umsetzen:

- Die Löschung von Daten im Internet muss möglichst einfach durchsetzbar sein.
- Die Nutzung von „Big Data“ müsste gesetzlich reglementiert werden (strengere Anonymisierungsvorschriften; insbesondere bei personenbezogenen Daten sollte nur eine repräsentative Menge genutzt werden, anstatt ganz große Datenmengen).
- Die Nutzungsbedingungen und Richtlinien von Internetdiensten und sozialen Netzwerken im Internet müssen erheblich vereinfacht und standardisiert werden.
- Es sollten mehr Server im eigenen Land vorhanden sein, damit die Datenströme nicht über die Server der USA oder Großbritannien laufen, die dort abgegriffen werden können. Politik könnte dafür die technische Infrastruktur fördern.
- Politik kann innerhalb einer Bildungsreform den oben angeführten Punkt fördern, dass der Umgang mit den technischen Möglichkeiten im Kontext des Internets erlernt und dementsprechend ein Pflichtfach in den Schulen werden sollte.

Werner Mittelstaedt, 24. März 2014

Literatur:

Bunz, Mercedes (2012): *Die stille Revolution. Wie Algorithmen Wissen, Arbeit, Öffentlichkeit und Politik verändern, ohne dabei viel Lärm zu machen.* Berlin: Suhrkamp Verlag.

Mayer-Schönberger, Viktor und Kenneth Cukier (2013): BIG DATA. Die Revolution, die unser Leben verändern wird. München: Redline Verlag.