

Das Problem des Datensammelns – einfach erklärt

Bei den Big-Data-Kolloquien zur Veranstaltung Freiheit 2.0 in Weil am Rhein wurde versucht, einem fachfremden Publikum die Problematik des Datensammelns durch Firmen bei der Nutzung von Onlineangeboten aus Sicht einer kritischen Informatik nahezubringen. Es wurde versucht, ein Problembewusstsein zu schaffen, indem die zugrundeliegenden Mechanismen der Informatik an einem verständlichen Gleichnis veranschaulicht wurden. Und es wurde beschrieben, wie Firmen die gesammelten Informationen im eigenen Interesse nach einer kommerziellen Logik, mitunter gegen die Interessen der Datenpreisgebenden einsetzen. Der folgende Text ist der Erklärungsversuch des Vortrags in überarbeiteter Fassung und soll vor allem als Orientierung für Informatiker:innen dienen, die selbst in die Situation kommen, das Thema einfach erklären zu wollen.

„Ich habe doch nichts zu verbergen“

Sucht man mit einer Bildersuchmaschine wie *Startuppage.de* nach „Überwachung“, findet man einige der Versuche, den Begriff der Überwachung symbolhaft darzustellen. Man findet Bilder von wachen Pupillen, Augen, die durch Schlüssellocher gucken, starrende Kameras auf Pfählen und an Häuserwänden, Ferngläser mit anonymen Beobachtern und auf Menschenköpfe gerichtete rote Fadenkreuze.



All diese Symbole und Metaphern bleiben in ihrer Darstellung jedoch bei dem Akt des Beobachtens stehen: Etwas Anonymes startet. Die benutzten Symbole selbst erzeugen auf die Betrachtenden den Effekt von Eingeschüchtertheit und Undurchschaubarkeit. Sie stellen also weder den Mechanismus der Einschüchterung, wie Jeremy Bentham ihn mit seinem Panopticon beschreibt, noch die hinter der Überwachung ablaufenden Prozesse der Informationsverarbeitung bildlich dar. In den Bildern wird man als Betrachter, in selbst der betroffene Mensch. Oft hört man als Reaktion auf die Konfrontation mit dem Thema Überwachung den Satz „Ich habe doch nichts zu verbergen.“ Dieser rührt genau aus dieser Perspektive her. Man nimmt sich zwar selbst als beobachtet wahr, doch antizipiert als den Zweck der Beobachtung, dass Fehlverhalten entdeckt und geahndet werden soll. Man nimmt an, die einzig negative Konsequenz aus der Überwachung für einen selbst sei die zu erwartende Reaktion auf eine unerwünschte Handlung, die beobachtet wird. Da man von sich selbst in diesem Zusammenhang annimmt, nicht gegen die antizipierten Regeln zu verstoßen und daher selbst nicht im Fokus der Überwachung zu stehen, erwartet man keine Ahndung und sieht kein Problem mit dem Beobachtetwerden. Grund zu dieser Annahme hat man jedoch nur, wenn man annimmt, dass Überwachungsmaßnahmen ausschließlich der Durchsetzung von Regeln oder einem Sicherheitsgewinn dienen sollen und gegen bestimmte Geschehnisse gerichtet sind.

Im Gegensatz zur Überwachung durch Kameras im öffentlichen Raum, wird bei der Sammlung von Daten durch Anbieter von Onlinediensten wie sozialen Netzwerken und Verkaufsplattformen selten der Begriff Überwachung benutzt.

Trotzdem scheinen deren User bei der Abwägung über die Preisgabe von Informationen einer ganz ähnlichen Logik zu folgen. Man nimmt an, aus dem, was man an Inhalten preisgibt, könne einem kein Strick gedreht werden, da man die Art der preisgegebenen Inhalte kontrollieren könne und nicht gegen Regeln verstoße. Diese Art der Reaktion auf Warnungen von Datenschützer:innen offenbart jedoch ein Missverständnis, das Edward Snowden in einem Satz auf den Punkt bringt: „Es geht nicht darum etwas zu verbergen, sondern etwas zu verlieren.“

Um zu verstehen, was es zu verlieren gibt, soll erklärt werden, nach welchen Prinzipien die gesammelten Daten ausgewertet und genutzt werden und welche Rolle die Informatik mit ihrer Herangehensweise dabei spielt.

Informatik sieht die Welt als Zahlen

Für Nichttechniker wird der Begriff Informatik zugänglicher, wenn man sie darauf aufmerksam macht, dass es im Kern nicht um Computer, diese unbegreifbaren Blechkisten mit dem Bildschirm geht, die andauernd nicht das tun, was die User wollen, sondern dass Informatik die Wissenschaft der automatisierten Verarbeitung von Informationen ist. Wichtig zu wissen ist allerdings, dass die Verarbeitung dabei aber sehr wohl den begrenzten Möglichkeiten einer solchen Blechkiste unterliegt. Dass die Hardware nämlich nichts anderes kann, als Zahlen zu speichern und nach formalen Regeln zu verarbeiten. Die Informationsverarbeitung unterliegt damit nicht nur den begrenzenden Regeln der Physik, sondern auch den Grenzen der Berechenbarkeit, an die man schneller stößt, als einem als Informatiker:in lieb ist.

Um also Informationen im Computer zu verarbeiten, müssen diese in abstrakter Form, d. h. als Zahlen darstellbar vorliegen. Um Informationen in diese Form zu bringen und dabei nicht am Speicherplatz oder der Langsamkeit der Berechnung zu scheitern, bedient sich die Informatik des Mittels der Abstraktion. Der Computer rechnet nicht mit den Sachverhalten selbst, um die es gehen soll, sondern mit einem mathematisch beschreibbaren Modell des Sachverhaltes. Ein solches Modell ist nicht nur eine einfachere Darstellung der Wirklichkeit. Die Einschränkungen, Weglassungen und Vereinfachungen, aus denen sich das



Geheimagent 2, Rudolf Schönwald, Kaltnadelradierung 1976, gemeinfrei

Modell dann ergibt, folgen bestimmten Annahmen, die sich aus dem Ziel ergeben, das mit der Informationsverarbeitung verfolgt wird.

Die Unterschiedlichkeit von Modellen ein und derselben Sache für verschiedene Zwecke wird gern anhand von Katzen erklärt: Denken unterschiedliche Menschen an eine Katze, haben sie unterschiedliche Modelle einer Katze im Kopf. Ein Katzenbesitzer, der seine Katze vor allem streichelt, hat ein Katzenmodell im Kopf, das eher einem weichen Fellball mit Augen, Schwanz und Beinen gleichkommt. Das Wesentliche eines Katzenmodells einer Tierchirurgin sind hingegen Skelett und Organe. Beide Modelle funktionieren für die jeweilige Aufgabe (Streicheln bzw. Operieren), sind dabei aber sehr unterschiedlich und nicht austauschbar.

Modelle sucht und benutzt die Informatik nicht nur für Informationen, die mathematisch leicht und nahezu abschließend beschrieben werden können, sondern auch für abstraktere und viel schwieriger und uneindeutiger mit mathematischen Mitteln beschreibbare Sachverhalte wie etwa die Interpretation eines Gesichtes in den Bildern eines Videos.

Ein anschauliches Beispiel für die Findung und Nutzung eines Modells ist der Wetterbericht: Jeden Tag aufs Neue sieht man in den Medien Landkarten, auf denen sich bunte Flächen bewegen, die für Temperatur, Wolken, Wind und Luftdruck stehen. Dies wird gezeigt, obwohl es keinen vollständigen aktuellen Datensatz über Luftdruck und Temperaturen für jeden Qua-

dratkilometer des gezeigten Gebietes gibt. An vielen kilometerweit auseinanderliegenden Messstationen werden einzelne Aspekte des Wetters gemessen und ausgewertet. Aus jahrelangen Beobachtungen wurden mathematische Modelle des Wetters erstellt, mit denen man anhand der aktuellen Messungen Prognosen berechnet. Obwohl nur einige wenige Messpunkte bekannt sind, werden für jeden einzelnen Punkt auf der gezeigten Karte Informationen gegeben, um zu entscheiden ob ein Regenschirm oder eher Sonnencreme nötig ist. Allerdings wird die Vorhersage für einen Ort umso ungenauer, je weiter die Prognose in die Zukunft reicht und je ungenauer das Modell eine geografische Besonderheit einer bestimmten Region berücksichtigt.

Vermessung der User

Vergleichbar mit der Messung von Wetterdaten durch einzelne verteilte Messstationen, werden Informationen erhoben, die Menschen bei der Nutzung von Internetdiensten wie sozialen Netzwerken, Online-Marktplätzen und/oder während der Benutzung ihrer Smartphones erzeugen. Ganz im Gegensatz zur Annahme, dass es hier nur um die wesentlich erstellten und geteilten Inhalte wie geschriebene und gesprochene Worte oder Fotos geht, werden bei der Nutzung wesentlich mehr Informationen erhoben. Um im Bild des Wetterberichtes zu bleiben – es werden unzählige von Messpunkten auf der ganz persönlichen Landkarte des datengebenden Menschen gesammelt. Die einzelnen Gebiete dieser persönlichen Landkarte stehen für verschiedenste Aspekte eines einzelnen Menschen, seines Lebens, seiner Handlungen und seiner Psyche. Weitläufige Gebiete dieser Landkarte der Aspekte unserer selbst sind bereits von einigen wenigen Firmen abgedeckt:

Amazon hat Messpunkte darüber, welche Bücher wir in welcher Geschwindigkeit lesen und wie viel wir für Technik zu bezahlen bereit sind. *Spotify* sammelt Informationen, zu welchen Zeiten wir welche Musik hören und kann Schlussfolgerungen über unsere Stimmung ziehen. *YouTube* und *Netflix* führen Listen darüber, wofür wir uns interessieren und welche Szenen wir uns zweimal anschauen. Die Spieleplattform *Steam* kann unsere Reaktionszeiten bei Simulationsspielen messen und unsere Problemlösestrategien bei Rätselspielen aufzeichnen. *Facebook* speichert, wann wer was liest und teilt uns mit, mit wem wer über was wann in welchen Kreisen kommuniziert. Fitness Apps messen und übermitteln unsere physische Konstitution und *Payback* schreibt mit, was wir essen, trinken, wie viel Benzin wir verbrauchen und welche Kleidung wir tragen.

Mit Sensoren in Smartphones für Luftdruck, Lage, Beschleunigung, Temperatur und hochqualitativen Mikrofonen wird jetzt schon den Apps unserer Telefone die Möglichkeit gegeben, viele der blinden Flecken auf unserer Landkarte zu beleuchten, die man mit der Art der Sensoren gar nicht intuitiv in Zusammenhang bringen würde. Forscher fanden zum Beispiel heraus, dass Bewegungen eines Schlafenden, die mit den Beschleunigungssensoren eines Smartphones, das nachts auf dem Bett liegt, gemessen wurden, mit hoher Wahrscheinlichkeit auf eine Disposition des Schlafenden für Alkoholismus geschlossen werden kann. Auf dem *Chaos Communication Congress 2016* schloss ein Datenforscher allein anhand der Veröffentlichungstermine

von Artikeln auf *Spiegel Online* auf die Urlaubszeiten der Autorinnen und Autoren und entdeckte, über einen längeren Zeitraum betrachtet, auch auffällig häufige Übereinstimmungen der Urlaubszeiten einzelner Autorinnen und Autoren. Beide Beispiele machen deutlich, was das Bundesverfassungsgericht schon beim Volkszählungsurteil 1983 feststellte: Es gibt keine mehr oder weniger schützenswerten Daten. Aus einzelnen, unbedeutend wirkenden Daten können Informationen privater Natur gewonnen werden. Jede Vorstellung, man habe bei der Nutzung von Onlineangeboten und ans Internet angebundener Hardware unter Kontrolle, welche Informationen man über sich selbst preisgibt, ist eine Illusion.

Komodifikation von Daten und Beeinflussbarkeit

Das Ziel einer kommerziellen Organisation, diese Mengen von Daten zu sammeln, ist deren Komodifikation – das Herauslesen auswertbarer Informationen und deren gewinnbringende Nutzung oder Weitergabe an Dritte. Das gewählte Modell, nach dem die Daten analysiert werden, ist genau wie das Katzenmodell entsprechend genau auf dieses Ziel ausgelegt. Aus den vielen einzelnen Messpunkten der vielen Gebiete unserer Landkarten wird kein möglichst vollständiger Blick auf uns als Mensch erzeugt, in dem wir uns wiedererkennen würden und mit dem wir uns identifizieren würden, sondern ein verzerrtes Abbild von uns entworfen, dessen Gestalt dieser kommerziellen Logik folgt. Es ist zu erwarten, dass diese Abbilder hauptsächlich dann gewinnbringend eingesetzt werden können, wenn man sie nutzt, um genau mit den Menschen Gewinn zu erzielen, von denen die Daten ursprünglich erhoben wurden. Der bereits offenkundig betriebene und einleuchtendste Weg, dies zu tun, ist das Einblenden individueller Werbung während der Nutzung der Dienste. Auch wenn die Ergebnisse zu den eigenen Interessen zu passen scheinen, muss man doch davon ausgehen, dass einem hier nicht die Produkte angeboten werden, die am besten zu den eigenen Ansprüchen, z.B. Haltbarkeit, Qualität und Reparierbarkeit passen, sondern Produkte von denjenigen Anbietern angezeigt werden, die für die Werbung bezahlt haben. In diesem Fall ist man selbst der- oder diejenige, der oder die das Geld durch den möglichen Kauf des Produktes wieder einbringt.

Weniger offenkundig sind allerdings die Möglichkeiten der Beeinflussung der Entscheidungen und Handlungen der Nutzen-

den – dem sogenannten Nudging (englisch anstupsen). Je mehr Informationen, je mehr Messpunkte auf der eigenen Landkarte man preisgibt, desto größer wird die Angriffsfläche für Beeinflussung und desto präziser können diese Beeinflussungen auf den Einzelnen und die Einzelne abgestimmt werden. Durch gezieltes Setzen oder Weglassen von Informationen kann die Beschäftigung mit Themen und Inhalten beeinflusst werden. Studien konnten außerdem zeigen, dass allein die Reihenfolge der Inhalte einer Timeline eines sozialen Netzwerks bereits signifikant die Diskussionen und Beschäftigung beeinflussen konnte. In diesem Fall bringt die Organisation, die ein Interesse an dem Einfluss hat, das Geld ins Spiel.

Neben diesen Arten von Beeinflussung werden, basierend auf dem gesammelten Wissen, Optionen eingeschränkt. Eine dafür besonders gewinnbringend verwertbare Information ist die über die angenommene Zahlungsfähigkeit. So wurden Kunden in Abhängigkeit der genutzten Hardwaremarke oder der ermittelten Region, aus der jemand die Verkaufsplattform in Anspruch nahm, unterschiedliche Preise und eine unterschiedliche Produktauswahl angezeigt. Es ist damit zu rechnen, dass in Zukunft eine ganze Palette derartiger Mechanismen der kommerziellen Ausnutzung der preisgegebenen Informationen genutzt werden wird.

In beiden Fällen – ob einerseits durch die Akzeptanz der Preisgabe von Daten und das Inkaufnehmen der Beeinflussung, oder andererseits durch die vorsichtige, stets überlegte Preisgabe von Daten und der damit verbundenen Einschränkung der Nutzung der Vorteile der Digitalisierung bis hin zum Ausschluss vom sozialen Leben, das sich zunehmend unter Nutzung datensammelnder Dienste abspielt – hat die Datensammlung einen Einfluss auf das eigene Leben, der nicht im eigenen Interesse liegt. Und nicht nur auf das eigene Leben – denn mit vielen Akten der Preisgabe von Daten gibt man auch Daten von anderen preis und trifft damit Entscheidungen für andere gleich mit.

Es geht bei der Kritik an Überwachung und Datensammelei nicht darum, etwas zu verbergen, was gegen die Regeln verstößt, sondern darum, nicht seine Selbstbestimmtheit und Gleichberechtigung zu verlieren und die von den Mitmenschen aufs Spiel zu setzen.

Twitter: [@algopticon](https://twitter.com/algopticon)

Website: www.algopticon.de



Benjamin Kees

Dipl.-Inf. **Benjamin Kees** studierte an der Humboldt-Universität zu Berlin Informatik und Psychologie. Schwerpunkte seines Studiums waren Überwachungstechnik am *Lehrstuhl für Informatik in Bildung und Gesellschaft* und Mensch-Technik-Interaktion am *Lehrstuhl für Ingenieurpsychologie*. Seit November 2015 ist Benjamin Vorstandsmitglied des FIF und Mitglied der Fachgruppe *Informatik und Ethik* der Gesellschaft für Informatik. Seit 2012 ist er IT-Leiter bei der *Smart Energy for Europe Platform* und seit Sommersemester dieses Jahres Lehrbeauftragter an der Hochschule für Technik und Wirtschaft Berlin.