

COPYRIGHT

Dieses Manuskript ist urheberrechtlich geschützt. Es darf ohne Genehmigung nicht verwertet werden. Insbesondere darf es nicht ganz oder teilweise oder in Auszügen abgeschrieben oder in sonstiger Weise vervielfältigt werden. Für Rundfunkzwecke darf das Manuskript nur mit Genehmigung von Deutschlandradio Kultur benutzt werden.

Deutschlandradio Kultur

AUS FORSCHUNG UND GESELLSCHAFT

BigData - die neue Kulturtechnik

ATMO

(Laborrauschen)

OTON (Englich) *Wei_Lab_BidDataCenter*

Dies ist unser Daten-Zentrum. Dies sind all die Maschinen, die wir in den letzten drei, vier Jahren gekauft haben. Sie arbeiten rund um die Uhr und generieren Big Data.

AUTOR

Der Biologe Wei Chen ist Leiter des Instituts medizinische System-Biologie am Max-Delbrück-Centrum in Berlin Buch. In einem mit Glaswänden abgetrennten Raum steht sein ganzer Stolz: Der Gensequenzierer, 4 Millionen Euro teuer und in nur einer Handvoll europäischer Labore zu finden. Ein Genom, den Bauplan eines Menschen, komplett zu

entschlüsseln - dafür hat das Human Genom Projekt Anfang der 90er xx noch 20 Jahre gebraucht.

OTON *Wei_Lab_Genom-in-wenigenTagen.WAV*

Mit dieser Maschine können wir das Genom innerhalb weniger Tage sequenzieren.

AUTOR

Die Konsequenz, so Wei: Bald dürfte die Technik so alltäglich sein wie ein Bluttest. Jedes Krankenhaus könnte komplette Genome analysieren und den individuellen Bauplan eines jeden Patienten lesen wie eine Bedienungsanleitung. Rund um den Globus würden Milliarden dieser Baupläne gespeichert - und Supercomputer könnten in diesem gigantischen Datenberg nach Mustern, neuen Erkenntnissen suchen.

OTON (Englisch) *Wei_Prognose_Krankheiten*

Anhand dieser Muster wird man sehen können, welche Menschen eine höhere Wahrscheinlichkeit haben, an bestimmten Krankheiten zu erkranken. Dann sind sie besser in der Lage, diesen Krankheiten vorzubeugen.

AUTOR

Die Welt erlebt eine Datenexplosion. Der Chip-Hersteller Intel rechnet damit, dass allein in diesem Jahr mehr Daten entstehen als in allen Jahren vor 2003. Alle Daten, die wir seit Beginn der Menschheit bis 2003 zusammengetragen haben - heute entstehe dieselbe Menge alle 2 Tage, sagt Google-Chef Eric Schmidt.

Nie haben Menschen so viele produziert, jede Sekunde: Nach dem Aufstehen, das Smartphone meldet sich beim Funkmast an; Mails checken,

Internet-Radio einschalten, Nachrichten lesen bringt Algorithmen auf Trab: in Millisekunden errechnen sie, wer vor dem Bildschirm sitzt, welche Anzeige mich interessieren soll und wer am meisten dafür bezahlt; bei Facebook die Kollegen auf Partyfotos markiert; unseren Weg zur Arbeit zeichnet das Mobilfunkunternehmen auf, Sensoren im Auto protokollieren Temperatur, Verschleiß und Verbrauch; das Navi sendet Geschwindigkeit und Ort an Staucomputer, in der Apotheke mit EC-Karte bezahlt und Datenbanken gefüttert, die meine Bonität dokumentieren - da ist der Tag erst zwei Stunden alt. Die digital vernetzte Gesellschaft sammelt Daten, erzeugt Daten, atmet Daten.

OTON Helbing_Datenexplosiom_Fakt

Das führt im Grunde dazu, dass wir in Zukunft sehr viele Daten sammeln werden. In der Gesellschaft, in den Unternehmen und so weiter. Man spricht vom Internet der Dinge: Sensoren, Prozessoren werden überall verteilt sein. In der Kaffeemaschine, im Kühlschrank, all das vernetzt sein mit unserem Handy. All das erzeugt natürlich Daten.

AUTOR

Der Informatiker Dirk Helbing erforscht an der ETH Zürich Potentiale und Gefahren von Big Data. Big Data meint: Nie gekannte Datenmassen, die in Echtzeit entstehen und von Computern in Echtzeit ausgewertet werden können. Nach Informationen, Mustern, Auffälligkeiten durchsucht werden alle Datenberge dieser Welt von Algorithmen, von Menschen geschriebene Computerprogramme, die funktionieren wie Filter: Such mir, lieber Google-Algorithmus, aus deinen Milliarden Webseiten jene zehn raus, von denen du glaubst, sie passten am besten zu meinen drei Suchworten. Diese Algorithmen durchsuchen auch Berge völlig neuartiger Daten: Nie haben Menschen so viele, so genaue und so aktuelle Spuren hinterlassen.

Mobilfunkdaten, Krankenakten, Einkäufe, Tweets bilden ein globales Nervensystem, das von uns unbemerkt sehr viel verrät über uns: Vorlieben, Verhalten, Pläne. Dirk Helbing.

OTON Helbing_Datenexplosiom_Fakt

Das ist ein industrieller Trend, der auf uns zukommt. Es ist nicht zu erwarten, dass sich daran etwas ändern - es sei denn, die Gesellschaft entscheidet sich, dass sie das nicht haben möchte. Aber es gibt sehr viele Potentiale und die Frage ist: Wie nutzen wir diese Daten?

AUTOR

Aus Mustern der Vergangenheit können Computer nämlich die Zukunft mitunter recht treffsicher vorhersagen: Revolutionen, Grippewellen und menschliche Handlungen - eine Vision, deren düstere Seite Steven Spielberg in „Minority Report“ verfilmte: Tom Cruise verhaftet Unschuldige, weil sie wahrscheinlich ein Verbrechen begehen. Daten wirken wie Naturkräfte: leise und verborgen, aber mächtig; sie bringen großen Nutzen und können erheblichen Schaden anrichten. Doch Daten sind keine Naturgewalt. Sie werden von Menschen gemacht und gesammelt.

OTON Helbing_Öl.wav

In Amerika beschäftigt man sich damit schon seit einigen Jahren. In Europa wacht man langsam auf, man sieht die Potentiale. Man spricht vom Öl des 21. Jahrhunderts, einer neuen Art von Industrie, die die Umwelt nicht verschmutzt, aber Risiken birgt, insbesondere für die Privatsphäre. Deswegen muss man sich sehr genau damit auseinandersetzen, welche Vorkehrungen getroffen werden müssen, damit die Informationsgesellschaft das Potential nutzt und die Risiken vermeidet.

AUTOR

Vor allem die Wissenschaft profitiert von BigData, der Möglichkeit riesige Datenmengen nach Mustern zur Durchsuchen.

ATMO

(Applaus, Vortrag)

OTON *(Englisch)*

Eines der Probleme, das wir seit 60 Jahren versuchen zu lösen, ist maschinelle Übersetzung. Die Kombination aus statistischen Verfahren und BigData hat in den letzten Jahren dazu geführt, dass wir einen viel besseren Job machen bei der Übersetzung von Webseiten und anderen Texten in andere Sprachen.

AUTOR

sagt Rick Rashid, Forscher bei Microsoft. Seit Jahrzehnten versuchen Forscher, Maschinen das Übersetzen beizubringen - ohne großen Erfolg. Dank BigData werden Supercomputer jetzt einfach mit Daten gefüttert: Jedes UNO-Dokument, jedes EU-Papier wird in eine Vielzahl von Sprachen übersetzt - identischer Text in vielen Sprachen. Die BigData-Rechner machen daraus schlicht riesige Wörterbücher, allerdings mit Phrasen, Halbsätzen und Metaphern. So ist es heute möglich, dass Microsoft-Forscher Rashid spricht und der Computer seinen Vortrag live ins Chinesische übersetzt.

OTON *(Englisch)*

Ich spreche Englisch und hoffentlich hören sie mich in meiner eigenen Stimme in Chinesisch sprechen *(Chinesisch)*. Es gibt hier noch viel zu tun

(Chinesisch). Aber die Technik ist vielversprechend *(Chinesisch)*. Wir glauben, dass diese Technik bald die Sprachbarrieren zwischen den Menschen niederreißen wird. *(Chinesisch)* Ich persönlich glaube, dies wird zu einer besseren Welt führen.

AUTOR

Die Möglichkeit menschlich nicht zu erfassende Textmassen zu durchkämmen und zu kategorisieren, kann Revolutionen vorhersagen - und hätte Bin Laden finden können, sagt Kalev H. Leetaru vom Institut für Computing in den Geisteswissenschaften an der University of Illinois. Der Wissenschaftler hat einen Supercomputer mit 100 Millionen Zeitungs-Artikel gefüttert: Alle Artikel der New York Times, chinesische Web-Zeitungen, australische Online-Medien - alles, was er digitalisiert auftreiben konnte. Er ließ seine Algorithmen dann nach Worten suchen, die eine Stimmung beschreiben wie „grausam“ und „grässlich“ oder „erfreulich“ und „schön“. So zeigte sich, dass der Ton aller Artikel, die über Ägypten geschrieben wurden, über die letzten Jahre immer negativer wurde - bis der Ton den tiefsten Punkt seit dem zweiten Weltkrieg erreichte - Januar 2011, wenige Tage vor der Revolution:

OTON *(Englisch)*

Das sagt uns nicht, der Aufstand wird nächsten Freitag um fünf nach fünf losbrechen. Aber es sagt uns, dass die Welt erkennt, dass es in Ägypten düsterer wird. Es ist wie ein Wetterbericht. Da heißt es: Morgen 70 Prozent Regenwahrscheinlichkeit. Vielleicht regnet es nicht, aber es lohnt sich, einen Regenschirm mitzunehmen. Denn die Voraussetzungen sind da. Sie brauchen immer noch einen Katalysator, um die Dinge anzustoßen, aber die Voraussetzungen deuten immer mehr auf Regen hin.

AUTOR

Weil die 100 Millionen Artikel dann schon mal im Rechner waren, hat Kalev Leetaru mit ihnen gespielt.

OTON *(Englisch)*

Was würde passieren, wenn ich mal „Bin Laden“ eintippe, die gesamte Berichterstattung der letzten Jahre nehme und daraus eine Karte mache? Dann bekommen sie das.

AUTOR

Kalev Leetaru zeigt eine Weltkarte. Aus der ganzen Welt führen Linien nach Nordpakistan.

OTON *(Englisch)*

Sie bekommen nicht die Stadt, in der er gefunden wurde, sie kriegen auch nicht sein Haus. Aber alles deutet auf einen Bereich von 200 Kilometern um das Haus herum, in dem er gefunden wurde.

AUTOR

Computer filtern Erkenntnisse aus Datenmassen, die kein Mensch erfassen kann - auch der in Berlin forschende Gentech-Biologe Wei Chen ist begeistert von den Möglichkeiten von BigData.

OTON *(Englisch)* *Wei_Biologie_System_TippingPoint*

Für die Biologie bricht eine neue Ära an. Denn Biologie war für sehr lange Zeit eine beschreibende Wissenschaft. Jetzt mit BigData entsteht eine quantitative Biologie, eine System-Biologie. Mit traditioneller Biologie haben sie immer nur Teile eines Organismus studiert. Doch sehr oft ist das Ganze in der Biologie mehr als nur die Summe seiner Teile. Sie müssen das

Ganze betrachten, um zu verstehen, wie das System funktioniert. Das war uns immer schon klar, aber jetzt haben wir mit BigData erstmals die Möglichkeit dazu. Wir können mathematische Modelle von Organismen bauen und verstehen wirklich, wie sie funktionieren.

OTON *Wei-Assi_ExperimenteAmRechner.WAV*

Es ist eine Veränderung dessen, was Wissenschaft macht, was biologische Forschung macht: von dem Im-Labor-stehen-und-Stoffe-zusammen-mixen hin zu „Wir gucken in die Daten, die es gibt“.

AUTOR

Sagt der Bioinformatiker, Andreas Gogol-Doering, Mitarbeiter von Wei Chen.

OTON

Wir müssen vielleicht gar nicht mehr ins Labor gehen, um Sachen heraus zu finden. Sondern wir schauen, was gibt es denn da für Daten und können wir aus den bereits bestehenden Daten etwas heraus ziehen, das so vielleicht noch niemand heraus gezogen. Ich muss nicht immer Experimente im Labor machen, ich kann auch Experimente im Computer machen. Ich kann mir Daten herunter ziehen, die es schon gibt aus dem Netz, riesige Datenmengen von Experimenten, die Leute vorher gemacht haben, und sie auf neue Aspekte befragen, an die noch niemand gedacht hat.

AUTOR

Hier begeben sich die Biologen in einen Bereich von Big Data, der das Durchforsten großer Datenmengen so viel versprechend macht - der mit deutschen Vorstellungen von Datenschutz aber schwer zu vereinbaren ist:

Sie kombinieren Datensätze, die für andere Zwecke erzeugt wurden. Nach deutschem Datenschutzrecht dürfen Daten nur für jenen Zweck gebraucht werden, für den sie erhoben wurden. Dieser Grundsatz jedoch ist mit BigData kaum zu vereinbaren. Denn richtig interessant wird es erst, wenn Forscher mehrere, zunächst unabhängige Datensätze kombinieren. Kaufhistorie von Online-Shops, Daten aus sozialen Netzwerken, Bewegungsprofile von Fluggesellschaften und Mobilfunkanbietern - aus Mustern der Vergangenheit errechnen Maschinen die wahrscheinliche Zukunft.

OTON Helbing_Vorhersage01

Im Grunde ist es so, dass Menschen voraussagbarer sind in ihrem Verhalten, als sie das gemeinhin so denken.

AUTOR

Dirk Helbing, Informatiker an der ETH Zürich.

OTON Helbing_Vorhersage02

Im Grund haben die Menschen einen relativ regelmäßigen Tageablauf, einen Wochenrhythmus. Wir gehen um eine bestimmte Zeit ins Büro, dann vielleicht ins Fitnessstudio, bringen die Kinder in die Kita und holen sie wieder ab. Dieser Rhythmus wird aus diesen Daten sehr genau sichtbar und noch vieles mehr, wer mit wem in Kontakt ist und so weiter. Das heißt, unsere Vorstellungen, dass soziale Prozesse unvorhersehbar sind, die muss man wahrscheinlich revidieren.

AUTOR

Dank BigData wissen Computer mitunter Dinge über uns, bevor wir sie wissen. Eine US-Handelskette liess Ihre Computer alle Einkäufe analysieren,

die Kunden gemacht hatten. Die Maschinen fanden heraus, dass eine Kundin schwanger ist - noch bevor diese es selber wusste.

OTON

Diese Episode ist eher so eine amüsante Geschichte, die aber zeigt, was mit BigData möglich ist.

AUTOR

Sagt der Datenwissenschaftler Benedikt Köhler. Ein Vater bekam Werbung von der Supermarktkette Target, in dem ihm und seiner Tochter Schwangerschaftsartikel ans Herz gelegt wurden. Der Mann empörte sich, das die Kette junge Menschen zur Schwangerschaft animiere und verlangte eine Entschuldigung. Doch dann kam seine Tochter und sagte, sie sei tatsächlich schwanger. Wie hatte die Werbemaschine das wissen können? Die Rechner durchforsten einfach riesige Datenberge zum Einkaufsverhalten der Kunden, erklärt der Daten Wissenschaftler Benedikt Köhler. Die Software sucht schlicht in der Vergangenheit nach Auffälligkeiten und stößt in den Zahlen dann etwa auf eine Tatsache wie:

OTON Köhler_Target

Jemand, der eine bestimmte Creme, einen bestimmten Puder kauft, da ist die Wahrscheinlichkeit sehr groß, dass diese Person sechs Monate später die ersten Windeln kauft. Also ein ganz banaler Zusammenhang zwischen zwei Warengruppen in dem Datensatz. Diese Zusammenhänge, wenn die hoch signifikant sind, dann kann ich die eben ausnutzen. Also wenn jemand diese Produkte kauft, weiß ich, diese Person ist im dritten Monat schwanger, weiß das in der Regel sogar, bevor die Person das selber weiß und kann das dann bewerben.

AUTOR

Textanalysen in Echtzeit, Schwangerschaftsprognosen und biografische Details aus Facebook-Likes - die Datenströme werden immer breiter, schneller und reichhaltiger. Sie umziehen unseren Globus wie ein weltweites Nervensystem.

OTON Helbing_Planetare_Nervensystem

Das planetare Nervensystem ist nun die Idee, dass wir diese Technologien für etwas Gutes nutzen können, um über den Zustand unserer Welt mehr zu erfahren - und zwar in Echtzeit. Und zwar nicht nur über unsere Umwelt, die physikalische und biologische Umwelt, wie man das heute schon macht, sondern letztendendes auch mehr lernen über das soziale Klima, über das politische Klima, über den Zustand der Märkte und der Wirtschaft, um auf diese Art und Weise Probleme zu vermeiden, die dadurch entstehen, dass wir entweder bestimmtes wichtiges Wissen nicht zur Verfügung haben oder dass es zu spät kommt und auf diese Weise es zu falschen Entscheidungen oder zu Systeminstabilitäten kommen kann.

AUTOR

Nicht nur private Unternehmen sind fasziniert von der Macht der Daten. Zu den größten Datensammeln gehören Staaten. Etwa im Rahmen der Vorratsdatenspeicherung. Die EU zwingt Telekommunikationsanbieter zu speichern: Wer hat mit wem wann wie lange telefoniert? Von wo hat er angerufen? Wem eine Mail geschrieben? All das muss mindestens 6 Monate gespeichert werden. Von allen Bundesbürgern. Mit diesen Informationen lassen ganze Gesellschaften analysieren, erläutert Daten-Wissenschaftler Michael Kreil am Beispiel der 22.000 Emails in seinem privaten Postfach:

OTON *Kreil_Mal-Netzwerk-Analyse*

Daraus habe ich mal ein soziales Netzwerk generiert. Also wann habe ich wem eine Email geschickt? Ich habe das grafisch dargestellt und es war überraschend, weil diese Grafik unglaublich gut mein soziales Umfeld dargestellt hat. Ich habe genau gesehen: Wer sind meine Geschäftskunden, wer sind meine engsten Freunde? Das ist für den einzelnen natürlich sehr interessant, aber man kann damit auch schlechtes machen. So gibt es Firmen, die verkaufen Hard- und Software, die analysieren alle Mails und alle Telefonate einer Bevölkerung beispielsweise.

AUTOR

Diese Firmen heißen etwa Gamma und Trovicor. Ihre Techniken werden nach Erkenntnissen der Menschenrechts-Organisation „Reporter ohne Grenzen“ in Ländern wie Syrien und dem Iran eingesetzt: Hochleistungsrechner und Software überwachen die komplette Kommunikation eines Landes, können in Telefonaten Stichwörter erkennen, Chats mitlesen und abbrechen, Web-Traffic nach Stichwörtern durchsuchen und unterbinden. Matthias Spielkamp, Vorstandsmitglied bei Reporter ohne Grenzen:

OTON

Wir werfen denen vor, dass sie daran beteiligt sind, Informationen zu sammeln für autoritäre und repressive Regimes, die dann eben einfach dazu führen, dass Dissidenten, das können Aktivisten, Journalisten oder Blogger sein, abgehört werden, inhaftiert werden, dass ihre Quellen auffliegen und dass Leute ins Gefängnis kommen und zum Teil gefoltert werden.

AUTOR

Auch demokratische Staaten nutzen BigData-Techniken, um ihre Bürger zu überwachen. In den USA sammelt die National Security Agency alle Daten, die sie kriegen kann. Um sie zu speichern und mit Algorithmen durchforsten zu können, baut die NSA gerade eines der weltweit größten Rechenzentren. Eine Vision hat längst viele ergriffen: Live-Daten aus allen Ecken der Gesellschaft; Supercomputer, die diese Informationen in Echtzeit verarbeiten, interpretieren und Vorhersagen treffen - mit dieser Technik könnten sich ganze Gesellschaften steuern lassen.

OTON *Helbing_Wohllollender-Diktator-no*

Wir haben da Bedenken.

AUTOR

Der Informatiker Dirk Helbing.

OTON *Helbing_Wohllollender-Diktator-no*

Das Konzept des Wohllollenden Diktators hat grundsätzliche Grenzen, die darin begründet sind, dass komplexe Systeme in Echtzeit nicht optimiert werden können. Das heißt, selbst wenn man die besten Absichten hätten, wäre es überhaupt gar nicht möglich, für alle die besten Entscheidungen zu treffen. Das ist in der Vergangenheit auch immer schief gegangen.

AUTOR

Private Unternehmen und staatliche Behörden horten also Datenmassen nie gekannten Umfangs. Doch Maschinen sammeln nicht mehr nur Datenschnipsel und verdichten sie zu Information. Immer öfter treffen Maschinen basierend auf diesen von ihnen selbst generierten Informationen auch Entscheidungen. Wie das geht, erforscht Michael Mai

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme:

OTON *Mai_Potentialanalyse06_Effizienz-Versicherungen*

Ein Beispiel kann sein, die der Prüfung von Schadensfällen in der Versicherung, wo sie klassisch sehr komplexe Prozesse habe, wo ein Sachbearbeiter drin sitzt. Wo sie aber auch mit Verfahren des Maschinellen Lernens 80 Prozent der Fälle automatisiert erschlagen können. Wo sie sagen, das ist ein Standardfall, das bewilligen wir so. Dann können sich die Sachbearbeiter auf die zehn Prozent der Fälle konzentrieren, wo noch menschliches Zutun nötig ist.

AUTOR

Firmen versprechen sich von solchen automatisierten Verfahren mehr Effizienz und weniger Kosten. Doch das Schicksal der betroffenen Menschen wird in die völlig neue Hände gelegt. Wer bekommt einen Bankkredit? Wer wird zum Vorstellungsgespräch eingeladen? Wer landet aufgrund seiner Facebook-Likes vielleicht mal auf Fahndungslisten?

OTON *Meinel_Maschinen-entscheiden*

Was das Neue ist, dass diese Schlüsse, die Verknüpfungen durch Maschinen, automatische Programme erstellt werden. Und die sind zunächst mal lange nicht so gut wie der Mensch.

AUTOR

Sagt der Informatiker Christoph Meinel, Leiter des Hasso Plattner Instituts der Uni Potsdam. Denn Maschinen werden von Menschen programmiert und können sich irren. „Richtige Entscheidung“, heißt bei Maschinen immer „mit 70, 80 oder 90 Prozent Wahrscheinlichkeit die richtige Entscheidung“. Es bleibt immer die Chance, falsch zu liegen, sagt Informatiker Meinel:

OTON *Meinel_Bessermachen*

Jedenfalls wird das gefährlicher sein, als wenn Menschen, die dann doch mit einem Hintergrundwissen solche Dinge leichter einordnen können.

AUTOR

Denn Algorithmen finden in den Datenbergen Muster. Ob sie die Realität abbilden, muss erst überprüft werden.

OTON *Wei-Assi_DatenQualität01*

Denn wenn sie lange genug suchen, finden Sie immer was.

AUTOR

Andreas Gogol-Doering, der Bioinformatiker im Labor von Wei Chen, nennt ein Beispiel, wie Maschinen aus Datenmassen scheinbar eindeutige, aber doch falsche Erkenntnisse destillieren:

OTON

Da gibt es Studien, die einen Zusammenhang zwischen Linkshänder-Sein und Einkommen belegen. Wenn sie Linkshänder sind, haben sie ein höheres Einkommen. Da ist natürlich niemand ran gegangen und gesagt, lass uns mal gucken, ob Linkshänder ein höheres Einkommen haben. Sondern sie haben eine riesige Menge Daten genommen und einfach nach Zusammenhängen gesucht. Eine Gefahr ist das, weil man sieht Zusammenhänge, die nicht da sind. Man denkt dann, ich habe hier einen Zusammenhang gefunden zwischen Linkshänderschaft und Einkommen - der existiert aber nicht, der war einfach zu zufällig in den Daten.

AUTOR

Um die Qualität dessen zu beurteilen, was die Maschinen aus den Datenbergen rausfiltern, sagt der Bio-Informatiker Andreas Gogol-Doering, müsse zunächst jeder verstehen, wie genau diese Filter funktionieren. Doch das weiß kaum jemand.

OTON *Wei-Assi_DatenQualität_Alogos_BlackBox*

Machine-Learning Algorithmen sind wie schwarze Kästen, wie Black Boxes, die dann Ergebnisse produzieren anhand von Daten und niemand versteht mehr im Einzelnen, warum die Maschine jetzt diese Entscheidung getroffen hat auf der Grundlage von diesen Daten. Da blickt keiner mehr durch.

AUTOR

Nach welchen Regeln die Maschinen Filtern und aufgrund welcher Faktoren sie dann Entscheidungen treffen, sind mitunter selbst jenen ein Rätsel, die diese Maschinen betreiben.

ATMO *Noller_Admin_Prozess_Durchlauf*

„Jetzt gehe ich mal auf einen Ad-Server rauf, davon haben wir derzeit 22...“

AUTOR

Im Büro der Firma Nugg.ad.

Nugg.ad hilft dabei, Werbeanzeigen im Web nur jenen zu zeigen, die sich dafür interessieren. Dazu kooperiert Nugg.ad mit Webseiten und speichert für jeden Surfer, was welche dieser Seiten er sich angesehen hat: War der Besucher zuletzt 20 Mal auf Sportseiten, zehn Mal auf Wirtschaftsseiten, ist der Mensch vor dem Rechner wahrscheinlich ein Mann und bekommt eher Autowerbung zu sehen.

OTON *Noller_Algo-keinMenschVersteht*

Das ist ein recht leicht verständliches Beispiel. Aber wie der Algorithmus rausfindet, ob man 20,29, 30 oder 29 ist, oder sich für ein spezielles Produkt interessiert, kann man nur noch schwer nachvollziehen als Mensch. Aber die Maschine kann halt sehr komplexe, multidimensionale Berechnungen machen und daraus ihre Schlüsse ziehen.

AUTOR

Was tun? Das Bundesverfassungsgericht hat entschieden: Jeder hat das Grundrecht auf informationelle Selbstbestimmung, meint: Bürger sollen wissen können, was Behörden und Firmen über sie wissen. Sonst könnten Bürger die Folgen ihres Handelns nicht mehr abschätzen und demokratische Teilhabe als zu riskant betrachten. Das Grundrecht auf digitale Intimsphäre betrachtet Smartphone und Computer als ausgelagertes Gehirn, auf das Staaten und Firmen nur in absoluten Ausnahmefällen zugreifen dürfen. Selbst der deutsche Staat missachtet diese Grundrechte, indem er die Rechner von Verdächtigen durch Spionage-Software ausforschen lässt. Wie sollen diese Rechte erst gegen global agierende Daten-Firmen durchgesetzt werden?

AUTOR

Darauf hat niemand eine Antwort. Wenn heute Datensätze wie Krankenakten, Mobilfunkdaten und Google-Suchen veröffentlicht werden, ist das Allheilmittel: Anonymisierung. Wenn alle Hinweise auf die Person wie Name oder Geburtsdatum entfernen - so das Kalkül - dann ist die Privatsphäre der Betroffenen geschützt. Doch das ist ein Irrtum, sagt der Schweizer Informatiker Dirk Helbing:

OTON Helbing_Anonymisierung-no

Anonymisierung ist im Grunde ziemlich wirkungslos. Sobald man mehrere

Datensätze verbindet, kann man im Grunde jedes Individuum identifizieren.

AUTOR

Denn jeder neue Datensatz verfeinert die Filter. Ein Beispiel: Datensatz Nummer 1 sind alle Urlaubsfotos vom Brandenburger Tor. Auf einigen ist ein unbekannter Mann zu sehen. Mit einer Gesichtserkennung lassen sich alle Fotos herausfiltern, auf denen dieser Mann zu sehen ist. Alle diese Fotos enthalten den sekundengenauen Aufnahmezeitpunkt und offenbaren ein deutliches Muster: Wann genau war der Unbekannte in den letzten zehn Jahren am Brandenburger Tor? Mit Datensatz Nummer 2, den Protokollen der Funkmasten am Brandenburger Tor, lässt sich nun filtern: Welche Telefonnummer war immer eingebucht, wenn der Unbekannte am Brandenburger Tor war? Jetzt fehlt nur noch Datensatz Nummer 3, der eine Verbindung herstellt zwischen Nummer und Name - ein Telefonbuch.

OTON *Helbing_Daten_Brauchen-kontrolle (OTON Helbing_Daten-Allgemeingut)*

Persönliche Daten sind unsere Daten, aber wir wissen nicht, was damit gemacht wird, das ist das Problem.

AUTOR

Der erste Schritt zur informationellen Selbstbestimmung ist die Befreiung aus der auch selbstverschuldeten Unmündigkeit. Wer nicht weiß, wo im Browser der Cookie-Schalter ist und was er bewirkt, hat sein Daten 1-Mal-1 nicht gelernt; wer nicht weiß, dass er Gratis-Dienste im Web mit seinen Daten bezahlt, sollte an Biofleisch denken: Qualität kostet. Doch um seine Währung wählen zu können - Daten oder Geld, Verzicht oder Teilhabe - brauchen Menschen Informationen.

OTON *Helbing_Loesung-Sorge-Missbrauch-Transaprenz*

Das Missbrauchspotential ist groß und ich glaube, die einzige Lösung, die es dafür gibt, ist Transparenz.

AUTOR

Sagt der Züricher Informatiker Dirk Helbing.

OTON

In erster Linie ist es wichtig, dass wir eine Einsicht in die Daten bekommen, die über uns gesammelt werden, dass wir entscheiden können, was mit diesen Daten gemacht werden darf, sie auch korrigieren und löschen können. Und das muss erst mal technologisch gelöst werden.

AUTOR

Mehr Kontrolle über unsere Daten - wie das in der Praxis aussehen könnte, beschreibt Michal Kosinski von der Universität Cambridge:

OTON (Englisch) *Michal_Kosinski_Privat-Daten-bei-uns*

Heute ist der akzeptierte Standard, dass alles, was Du tust, irgendwo im Netz gespeichert wird. Welche Bücher wir kaufen, welche Websites wir besuchen, welche Suchanfragen wir stellen - all das wird bei unterschiedlichen Anbietern gespeichert.

AUTOR

Das müsse nicht so sein, sagt Cambridge-Forscher Michal Kosinski:

OTON (Englisch) *Michal_Kosinski_Privat-Daten-bei-uns*

Man kann sich vorstellen, dass diese Informationen explizit auf unseren eigenen Rechner gespeichert werden oder in einem privaten Datensafe im

Netz. Diese Daten sind sehr wertvoll und nützlich. Man kann Suchergebnisse verbessern, Vorhersagen und Empfehlungen ableiten: Welches Buch könnte dich interessieren? Aber all diese Berechnungen sollten aber auf unserem eigenen Rechner passieren. Und dann entscheiden wir, ob wir die Ergebnisse den Web-Firmen zeigen oder nicht.

AUTOR

Transparenz ist nicht nur für die Daten notwendig, sondern auch für die Regeln, nach denen sie gefiltert, bearbeitet und durchforstet werden, sagt Informatiker Helbing.

OTON Helbing

Heutzutage bekommen wir irgendwelche Empfehlungen oder Suchresultate. Wir wissen aber nicht auf welcher Basis, welcher Qualität die Daten sind, wie stark wir manipuliert werden und so weiter.

AUTOR

Helbing fordert, auch die Algorithmen offen zu legen, damit nachvollziehbar wird, wie Maschinen unsere Wahrnehmung beeinflussen:

OTON Helbing_Loesung_Individuum_waehlt-filter

Wir sprechen da von einem Innovations- und Informations-Ökosystem, wo jeder selber seine Empfehlungsfiler selber auswählt und anpasst, an das, was gut für ihn ist, und mit seinen Freunden und Kollegen tauscht, so dass auf diese Weise immer bessere Informationsfilter entstehen - die aber nicht vorgeben sind von einer bestimmten Firma.

OTON Noller_Filter-Transaprenz

So faszinierend Big Data und Algorithmen sind, so groß das

Innovationspotential ist, glaube ich, dass die Gefahr auch ist, dass uns als Gesellschaft da was entgleitet.

AUTOR

Sagt Stephan Noller von Nugg.ad, jener Firma, die Algorithmen benutzt, um uns passende Werbung zu zeigen.

OTON *Noller_Filter-Transaprenz*

Algorithmen filtern, treffen autonome Entscheidungen, dass bestimmte Leute, bestimmte Artikel nicht mehr lesen sollen, weil der Algorithmus denkt, musst Du nicht. Finde ich vielleicht toll, weil ich Sportberichte nicht mag, aber wenn der Algorithmus entscheidet, dass ich über irgendeinen Politiker was nicht lesen soll, dann weiß ich nicht, ob ich das will. Ich finde es tatsächlich gesellschaftlich problematisch, wenn wir Filter-Infrastrukturen zulassen, ohne irgendwie einzufordern, dass es Hinweise gibt, dass es Abschaltmöglichkeiten gibt, dass es transparent gemacht wird. Das finde ich persönlich problematisch.

AUTOR

Menschen sollen über ihre Daten verfügen und die Regeln bestimmen, nach denen sie ausgewertet werden. Das bedeutet Veränderung auf allen Ebenen der Gesellschaft: Staaten und Unternehmen müssen Daten, den wertvollsten Rohstoff der digitalen Gesellschaft, gleicher verteilen, Bürgern Zugang gewähren und Wissen teilen. Doch auch die Bürger können nicht weiter machen wie bisher, müssen Ihr Verhältnis zu Daten und Algorithmen ändern: Sie müssen die passive Rolle derer verlassen, die nur Objekt staatlicher und privater Datenjongleure sind. Bürger müssen ihre Daten aktiv zu Eigen machen: Kinder müssen programmieren lernen, weil Algorithmen nur verändern kann, wer sie versteht; Menschen müssen

wissen wollen, wo im Browser sich Cookies sperren lassen und wie man einen Film schneidet. Viel Arbeit, sicher, gesteht auch Informatiker Helbing. Aber Daten sind eine neue Kulturtechnik:

OTON *Helbing_Daten-Allgemeingut-Lesen-Schreiben*

Nehmen wir einfach mal die Situation als Schreiben und Lesen noch ein Privileg war. Natürlich war das gut für die Leute, die Lesen und Schreiben konnten, denn sie hatten einen Vorteil. Und trotzdem hat sich die Gesellschaft entschieden, das Lesen und Schreiben allen Menschen beizubringen. Und das hat einen riesigen Schub für die Gesellschaft gebracht; das hat uns in der Geschichte nach vorne gebeamt. Und das gleich kann passieren, wenn wir die Daten für jedermann nutzbar machen.