Alles, was recht ist

Der richtige Umgang mit Daten

Bernd Fuhlert

Daten sind zunehmend leichter erhältlich. Anwendungen lassen sich so programmieren, dass sie möglichst viel Wissen über den Nutzer sammeln und weitergeben – alles für eine gute Analyse. Doch im Hinblick auf den richtigen Umgang mit personenbezogenen Daten sind einige Fragen noch nicht endgültig geklärt und andere bedenkenswert.

Individuelle, personenbezogene Informationen

Theoretisch ist heute sehr viel realisierbar – große Datenmengen können nicht nur generiert, sondern auch, exakt abgestimmt auf den Bedarf, analysiert werden. Insbesondere personenbezogene Informationen stellen dabei einen enormen Wert für Unternehmen dar, denn deren Auswertungen führen unter anderem dazu, dass diese neue (noch intimere) Kenntnisse über ihre Zielgruppe – und zwar speziell heruntergebrochen auf das Individuum – erhalten, um dann jeden so gezielt als möglich anzusprechen.

Dieser Grad der Individualisierung wird als eine der großen Chancen der Digitalisierung angesehen. Doch lässt sich diese betrachten, ohne dazu auch kritische Fragen zu stellen? Etwa, ob die Zusammenführung aller erfassbaren Daten, die auf den verschiedenen Ebenen eruiert werden, durchweg und für alle beteiligten Parteien nur positiv ist. Und dieser Gedanke dann weitergedacht: Kann die Zusammenführung großer Datenmengen auch negative Implikationen mit sich bringen? Diese Sachverhalte gilt es, zeitnah unter den verschiedenen Aspekten einer insgesamt verträglichen Nutzung zu untersuchen – auch weil die Technologie, aufgrund ihrer rasanten Entwicklung, mittlerweile nahezu alle Bereiche des täglichen Lebens durchdringt und auch beständig neue Daten entstehen.

Nicht zuletzt fällt diese Aufgabe auch Unternehmen zu, deren Geschäftsgrundlage und Weiterentwicklung zunehmend auf der Digitalisierung basieren: Hier bedarf es des Aufbaus einer Kompetenz dahin gehend, die Risiken technisch bezüglich Datensicherheit sowie rechtlich im Hinblick auf den Datenschutz einschätzen zu können und in der Umsetzung mit praxistauglichen Lösungen die Handhabbarkeit zu gewährleisten.

Diese Maßgaben gelten ebenso verbindlich für die Entwickler: Von daher sollte in der Entwicklungsumgebung unter anderem eines der vorrangigen Ziele sein, mit den richtigen Maßnahmen ein hohes Sicherheitsniveau zu etablieren – der Ansatz „Security by Obscurity" greift hier nicht, da die Analyse von Java-Programmen leicht realisierbar ist.

Was kann konkret passieren?

Hier gibt es vielfältige Aspekte: Aufgrund der Komplexität steigt auch die Vulnerabilität von Unternehmen. Mittlerweile können durch Datenpannen und Hackerangriffe mit Datenverlusten, Wirtschaftsspionage oder auch das Ausspähen umfangreich verknüpfter sensibler Persönlichkeitsdaten weitaus größere Schäden angerichtet werden, als dies bislang der Fall war. Hierzu gibt es bereits konkrete Vorfälle: zum Beispiel, dass große Mengen personenbezogener Daten unbefugt aus dem Unternehmen heraus an zahlende Auftraggeber weitergeleitet werden, etwa von frustrierten Mitarbeitern, die unbeschränkt Zugriff auf Kundendatenbanken haben. Dies kann – wie im Fall von Verizon – sogar über mehrere Jahre unbemerkt geschehen.

Aber, aus der Komplexität, gepaart mit dem Vertrauen in die Aussagekraft der Analysen, können noch ganz andere gravierende Folgen resultieren. Zum Beispiel durch die Fehlinterpretation bei der Bewertung der Daten. Denn zunehmend wird darauf gesetzt, durch die algorithmische Auswertung von großen Datenmengen Entscheidungen zu treffen oder Ereignisse zu prognostizieren, beispielsweise, um entsprechende Maßnahmen einzuleiten.

In den meisten Fällen sind diese Analysen jedoch überhaupt nicht oder zumindest nicht in Echtzeit nachvollziehbar. Dies macht es nahezu unmöglich vorauszusagen, ob der angewandte Algorithmus in einem Bereich liegt, in dem die Unschärfe der verwendeten Daten nur eine geringe Auswirkung hat, oder ob die Effekte gravierender sind. Wie aktuell geschehen: „Zum zweiten Mal innerhalb von nur fünf Monaten werden die Kapitalmärkte vollkommen auf dem falschen Fuß erwischt, weil scheinbar ausgefeilte Umfragen und Prognosemodelle, die mit Erfahrungen der Vergangenheit kalibriert wurden, den Wunsch nach einem politischen Wechsel und schlicht den Protest der Wähler nicht zutreffend abgebildet haben", zitiert das „Handelsblatt" Stefan Bielmeier, Chefvolkswirt der DZ Bank (Quelle: `http://www.pressetext.com/news/20161109016`).

Ist die digitale Welt überhaupt noch für jedermann verständlich?

Ein Beispiel: Bei der ständigen Suche nach Optimierung spielen gerade „Bots" eine große Rolle. Das sind Programme, die in der Regel mittels Sprach- oder Texterkennung automatisch auf einfache Anfragen des Nutzers antworten oder triviale Aufgaben wie die Einstellung der Weckzeit automatisch ausführen. Wenn es bei der Erledigung von Alltagsdingen bleibt, dann ist dagegen nicht unbedingt etwas einzuwenden.

Auf einer anderen Ebene – zum Beispiel in der politischen Kommunikation – kann der Einsatz von Bots hingegen gefährlich werden. Beispielsweise unter dem Aspekt der Manipulation. Ein solcher Sachverhalt lässt sich gut anhand des Wahlkampfs von Donald Trump illustrieren: Dieser wurde zu großen Teilen auf Twitter geführt – er selbst bezeichnete den Dienst in einem Interview als „seine persönliche Zeitung".

Unter der Annahme, dass nicht jeder, der Twitter oder andere soziale Netzwerke nutzt, auch ausreichend über deren Wirkungsweise aufgeklärt ist, liegt die Befürchtung nah, Nutzer halten alles uneingeschränkt für wahr, was dort zu lesen ist. Dies trifft jedoch keinesfalls zu – so sollen 37 Prozent der Tweets, die während eines Fernsehduells Trump zustimmten, nicht echt gewesen sein. Hinzu kommt, dass bei dem Einsatz von „Bots" Meldungen häufiger erscheinen, weil diese automatisch einen Retweet sowie eine Kommentierung auslösen. Mittlerweile wird jene Art der Online-Kommunikation sogar als Dienstleistung angeboten.

Wie schlecht kontrollier- beziehungsweise leicht manipulierbar „Bots" sein können, dafür hat Microsoft den Beweis geliefert – mit der Einführung von „Tay". Dieses selbstlernende Chat-Programm auf Basis von künstlicher Intelligenz musste schon nach kurzer Zeit wieder aus dem Netz genommen werden, weil Nutzer es geschafft hatten, das Programm mit rassistischen Kommentaren zu infiltrieren. Diese wurden dann ständig weiter getwittert – ohne Chance, dem Einhalt zu gebieten.

Abb. 1: Daten werden beständig produziert
(Quelle: #126852519 | © sdecoret - Fotolia.com)

Regelungen für den Umgang mit Algorithmen und Daten notwendig?

Vor einigen Monaten beklagten US-Verlage nach erneuten Anpassungen des Newsfeed-Algorithmus durch Facebook einen massiven – teilweise zweistelligen – Einbruch der Zugriffszahlen. Denn Posts von Freunden wurden nun höher gerankt und gleichzeitig die Beiträge von Pages im Newsfeed zurückgestuft. Potenziell ist davon auszugehen, dass nicht die Relevanz der einzelnen Nachricht im Vordergrund steht, sondern das Geschäftsmodell des Plattformbetreibers – und dies lautet: möglichst viele Werbeeinnahmen über gesponserte Posts.

Doch nicht nur unter diesem, sondern auch unter einem weiteren Aspekt gilt es einige Geschäftsmodelle, die momentan angedacht werden, zu beleuchten – vorrangig auch, dass so der „gläserne Mensch" entsteht, dessen Daten ausge- und verwertet werden, ohne dass er direkt darauf Einfluss nehmen kann: Zum Beispiel am Fall des Telekommunikationsunternehmens, das vorhandene Bestandsdaten (Geschlecht, Alter) seiner Kunden mit Bewegungsdaten verknüpft hat, die sich aus der Nutzung mobiler Endgeräte ergeben. Diese Daten sollten dann an Unternehmenskunden zur zielgerichteten Werbung weiterverkauft werden.

Ob im Bezug auf die Verwendung der Daten zukünftig neue Vorschriften geschaffen werden müssen, wird sich zeigen. Ein gewisses Maß an Regulierung bietet hier bereits die EU-Datenschutz-Grundverordnung (DS-GVO). Gemäß Artikel 5 Abs. 1 (b) der DS-GVO müssen personenbezogene Daten für genau festgelegte, eindeutige und rechtmäßige Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden. Der Umgang mit personenbezogenen Daten bleibt auch weiterhin verboten, wenn er nicht entweder durch einen Erlaubnistatbestand der DS-GVO oder einer sonstigen Rechtsvorschrift, beispielsweise einer Spezialgesetzgebung wie dem Telekommunikationsgesetz (TKG) oder Telemediengesetz (TMG) erlaubt ist – ein Grundprinzip lautet: „Verbot mit Erlaubnisvorbehalt". Wie hingegen der Umgang mit dem Einsatz von Algorithmen geregelt werden soll, wurde bislang nur vereinzelt diskutiert.

Ist Datensparsamkeit tatsächlich nur ein Hemmschuh?

Mittlerweile werden Daten als das Öl des 21. Jahrhunderts bezeichnet – ein Rohstoff, der das Überleben von Unternehmen sichern soll. Das Erwerben von Daten scheint so wichtig geworden zu sein, dass alle Interessen dem stückweise untergeordnet werden.

Die Kehrseite der Medaille könnte jedoch sein, dass aufgrund der – für jedes Unternehmen, das über die finanziellen Mittel verfügt – beliebig zugänglichen Daten in naher Zukunft eines auf der Strecke bleiben könnte: der freie Wettbewerb. Denn in der Vergangenheit konnte sich ein Unternehmen darüber differenzieren, dass es eigene Marketingdaten und exklusive Kenntnisse über die Zielgruppe im speziellen Markt hatte. Heute sind Daten in den verschiedensten Auswertungskonstellationen – zunehmend auch gebündelt in der Hand von externen Dienstleistern – verfügbar. Folglich bilden (theoretisch) in jedem Konzern die gleichen Daten, welche entsprechend gleich aufbereitet sind, die Basis für Entscheidungen.

Hier müsste zudem noch eine weitere Frage zur Diskussion gestellt werden: Können Daten überhaupt über einen längeren Zeitraum so sauber erfasst werden, dass die Messwerte gleichbleibend sind und nicht etwa die Aussagekraft der Analyse verfälscht wird, zum Beispiel durch das Hinzufügen von Merkmalen.

Lassen sich zentrale Fragen klären?

Dies zu klären, könnte für Unternehmen relevant werden. Denn momentan ist nicht vorhersehbar, ob und wann ein Umdenken bei den Konsumenten stattfindet, da Themen wie Big Data inklusive Auswirkungen der gegebenen Analysemöglichkeiten zunehmend in der öffentlichen Diskussion stehen.

Zudem gibt es bereits Gegenbewegungen, so wird momentan an Entwicklungen gearbeitet, die die Privatsphäre der Menschen wieder mehr schützen sollen: etwa die konkreten Pläne der deutschen Telekom, Unternehmen noch in diesem Jahr „Schutz vor Ausspähung und Angriffen durch Hobby-Drohnen anzubieten", oder die Erfindung des Alu-Mantels, der den Träger davor schützen soll, dass beispielsweise die Daten seines Mobiltelefons auf der Straße ausgelesen werden können.

Wie reagieren Kunden auf die zunehmenden Analysemöglichkeiten?

Viele Versicherer träumen ja schon davon, zukünftig Tarife exakt auf die Lebensumstände abzustimmen. Zum Beispiel kann durch den Einsatz von Fitnessarmbändern gemessen werden, ob der Versicherte sich auch ausreichend bewegt. Falls nicht, muss wegen der ungesünderen Lebensweise eben mehr bezahlt werden.

Das könnte dazu führen, dass irgendwann ein Schutz der Privatsphäre, etwa durch Datensparsamkeit, nur noch ein Privileg der Besserverdienenden ist. Wie werden Kunden darauf reagieren? Und was wird passieren, wenn sich aus den kumulierten Daten aller Versicherten neue Zusatzleistungen entwickeln lassen, die dann wiederum bei einem bestimmten Lebensstil obligatorisch werden?

Wem gehören die Daten?

Um die Frage zu beantworten, muss vorab geklärt werden, ob sich Dateneigentum überhaupt definieren lässt. Auf Basis des Datenschutzrechts ist dies nicht möglich, gemäß diesem ist der Begriff nicht existent. Fassbar ist er allenfalls im Kontext von Geschäftsgeheimnissen für Daten, die konkret auf einer Festplatte abgespeichert sind. Für Informationen über eine Person, so wie sie beispielsweise im vernetzten Auto entstehen, fehlt es an einem „materiellen Aufhänger".

Hierzu ein Beispiel: Werden Daten originär im Auto durch die Initiative des Automobilherstellers erzeugt, dann jedoch auf vielfältige Weise an verschiedenen Orten zu verschiedenen Zwecken verarbeitet, ist schwer zu definieren, wer letztendlich welche Daten wie verwenden darf.

Wer legt die Regeln für die Entwickler fest?

Die Maßgabe, dass das Vertrauen in die Analysemöglichkeiten allgemein ständig in der Diskussion steht, wirft die Frage auf, ob nicht einheitliche Vorgaben geschaffen werden müssen, die – auch unter ethischen – Gesichtspunkten gewährleisten, dass die Vorstellungen und Werte der Programmierer von Algorithmen keinen wesentlichen Einfluss auf das Ergebnis haben.

Zum Beispiel für die Berechnung von Kreditvorgaben: Müsste hier nicht für alle Beteiligten ersichtlich sein, welche Kriterien führen zu validen Ergebnissen und wie ist sichergestellt, dass diese auch tatsächlich Berücksichtigung finden?

Wie verlässlich ist die physische Welt?

Aufgrund der großen Mengen werden zunehmend Daten in die Cloud ausgelagert und auch die Intelligenz zur Auswertung. Das macht die Unternehmen abhängig von den Cloud-Anbietern. Falls hier keine geeigneten Konzepte zur Sicherung der eigenen Daten seitens der Unternehmen bestehen, könnte ein Angriff auf einen zentralen Anbieter wie Amazon Web Services (AWS) einen enormen wirtschaftlichen Schaden anrichten.

Das sollte Anlass genug sein, sich mit der Frage zu beschäftigen, ob die großen Cloud-Anbieter mittlerweile ebenso systemrelevant sind wie Banken. Denn ein Angriff auf einen der bekannten Provider hätte unmittelbar zur Folge, dass gleich Tausende von Unternehmen davon betroffen wären. Eine gewisse Zurückhaltung gegenüber diesen Dienstleistern ist somit angebracht.

Auch der zunehmenden Komplexität von Infrastrukturen muss Rechnung getragen werden. Um die Folgen, die aus kriminellen Angriffen resultieren können, zu minimieren, gilt es hier, die richtigen Schutzmaßnahmen zu etablieren: Dazu gehört einerseits, bereits bei der Entwicklung von Software allgemein den Ansatz „Secure Coding" zu verfolgen. Ebenso selbstverständlich sollte es sein, sensible Daten nur verschlüsselt zu übertragen sowie abzuspeichern.

Sicherheit muss in der Entwicklungsumgebung eine hohe Priorität haben

Die Tatsache, dass die Analyse von Java-Programmen leicht durchführbar ist, macht eine entsprechende Absicherung der Entwicklungsumgebung unerlässlich. Folgende Kriterien sind dabei relevant:

1. Kriterium: Grundsätzlich immer die aktuelle Java-Version verwenden.

2. Kriterium: Bei der Auswahl und dem Einsatz der Frameworks ist es wichtig, darauf zu achten, dass ...

- ... eine adäquate Verschlüsselung,
- ... sichere Datenbankzugriffe und
- ... eine sichere Authentifizierung realisierbar sind.

3. Kriterium: Den Schutz der eingesetzten Applikationsserver durch ...

- ... das Absichern der Schnittstellenkommunikation (z. B. SOAP, Corba, RPC-XML, REST, RMI),
- ... das Absichern der Namespaces,
- ... verhindern von Remote-Objekten,
- ... sichere Kommunikationskanäle in Multi-Tier-Umgebungen zu gewährleisten.

4. Kriterium: Organisatorisch – Definition des Rechte-Managements, zum Beispiel in Policies, um hohe Rechte-Anforderungen zu vermeiden.

Was ist für Unternehmen essenziell?

Bei dem Umgang mit personenbezogenen Daten gilt es für die Unternehmen, einiges zu beachten. Im Grundsatz sollten diese Daten nur dann erhoben werden, wenn es unbedingt erforderlich ist. Unter der DS-GVO muss das Unternehmen neuerdings den Betroffenen in der Datenschutzerklärung darüber informieren, auf welche Rechtsgrundlage man die Datenverarbeitung stützt. Falls dies unter Art. 6 Abs. 1 lit. f DS-GVO erfolgt, müssen auch die berechtigten Interessen des Verantwortlichen aufgeführt werden.

Versteckt normiert die DS-GVO Artikel 6 lit. f., dass die Datenverarbeitung immer dann zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich ist, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen.

In anderen Worten: Wenn ein Unternehmen ein berechtigtes Interesse, wie etwa den eigenen Geschäftszweck, an einer Datenverarbeitung vorweisen und zeitgleich der Nutzer nicht eine Schutzbedürftigkeit hinsichtlich dieser Daten geltend machen kann, dann ist die Verarbeitung zulässig. Diese Interessenabwägung findet sich auch im aktuellen Datenschutzrecht in ähnlicher Form wieder: zum Beispiel in § 28 Abs. 1 Satz 1 Nr. 2 BDSG (`https://www.gesetze-im-internet.de/bdsg_1990/__28.html`) – besser bekannt unter „Datenerhebung und -speicherung für eigene Geschäftszwecke".

Unternehmen, die Big-Data-Analysen zugeneigt sind, dürfen sich freuen. Allerdings nur eingeschränkt. Denn es kommt selbstverständlich nicht auf die Sicht des Unternehmens an, wann das eigene Interesse über die Interessen des Betroffenen zu stellen ist, sondern wie dieses Merkmal von den Datenschutzaufsichtsbehörden und anschließend den Gerichten ausgelegt wird.

Natürlich lässt sich durch Anonymisierung und Pseudonymisieren im Bezug auf die Verarbeitung von Daten der Handlungsspielraum vergrößern. Hierbei gilt allerdings zu beachten, dass die mehr erfordert als „den Namen des Nutzers aus dem Schlüsselfeld zu streichen".

Plädoyer: Mehr Sorgfalt im Umgang mit Daten

Diese Diskussion soll keinesfalls als Ablehnung der Technologie verstanden werden, sondern lediglich als Plädoyer für einen vernünftigen Umgang damit – entsprechend der kontinuierlich geführten Debatten bezüglich der Gefahren beim Klonen oder der Kernkraft. In Anlehnung daran gilt es, unter anderem zu erörtern, wie der Spagat zwischen dem Nutzen einer moderneren Lebensweise einerseits und einer, am Ende fehlinterpretierten, Experimentierfreudigkeit sowie zu hoher Risikobereitschaft andererseits gelingen kann.

Hier müssen allgemein anerkannte Grenzen festgelegt werden, nicht zuletzt um die Technik so einzusetzen, dass sie die Möglichkeiten der Gesellschaft erweitert, ohne den persönlichen Einfluss der Individuen zu verringern. Denn die Erwartung, dass durch große Datensammlungen die Realität komplett abgebildet werden kann, erinnert ein wenig an das Höhlengleichnis von Platon und führt zu der Frage: Entsteht eventuell hier nicht nur die Projektion der Wirklichkeit?

Daraus lässt sich folgern: Was letztendlich benötigt wird, ist mehr Transparenz – in erster Linie darüber, wie Verfahren, Strukturen und Entscheidungsprozesse funktionieren. Denn nur so lässt sich Kontrolle für alle beteiligten Parteien auch realisieren.

Bernd Fuhlert war als freier Dozent unter anderem an der Universität Duisburg/Essen, an der FOM Hochschule sowie an weiteren renommierten Institutionen tätig. Er verfügt über TÜV-Zertifizierungen als Datenschutzbeauftragter, Datenschutzauditor sowie als Chief Information Security Officer, der mit allen Fragen rund um § 11 BDSG-Audits (Bundesdatenschutzgesetz) vertraut ist. Seit 2016 ist er Geschäftsführer bei der @-yet GmbH. E-Mail: bernd.fuhlert@add-yet.de