Herausforderungen durch Big Data und Mustererkennung - Datensensibilität stärken

Wird durch Big Data und Mustererkennung das bislang Unsichtbare einer Gesellschaft sichtbar mit der Folge des Zugewinns an Erkenntnissen oder aber der Zerstörung der sensiblen Membran von Welt und Privatheit? Was wird überwiegen?

Unsere Perspektive ist, die wir hier erläutern wollen, es können nicht alle Daten „über einen Leisten geschlagen werden". Um die Datensensibilität zu stärken, ist ein differenzierter Blick auf die unterschiedlichen sozialen Datentypen zu werfen. Wir nehmen drei unterschiedliche Perspektiven ein.

Personenbezogene Daten

In dieser Perspektive werden die privaten Verhaltensdaten, die im Netz als Such-, Bewegungs-, Text-, Sprechdaten und vieles mehr anfallen, durch Mustererkennung in ihre wertvollen Substanzen zerlegt. Dahinter steht die Jagd auf das besondere Profil jedes Einzelnen, mit dem Ziel des Nutzer-Profilings. Sind die „Bestandteile" transparent, steht die Person mit Unterstützung von Künstlicher Intelligenz in ihrer „Besonderheit", den persönlichen Neigungen, der Gemütsverfassung oder der individuellen ökonomischen Situation „bloß da", sie ist vorhersagbar.

Das Ziel dieser Predictive Analytics ist: Mit Deep-Learning-Methoden die Einzigartigkeit (Singularität) jedes Einzelnen zu erkennen, um das Verhalten von Individuen schließlich überwachen und steuern zu können.

Das Netz „dringt nun von außen in die Privatsphäre ein – wo es nichts zu suchen hat, aber viel zu finden gibt", so der Sozialwissenschaftler Armin Nassehi. Damit rücke die Zerstörung der unsichtbaren Membran von Welt und Privatheit, in der wir geschützt leben wollen, in den Fokus. Big Data kombiniere Daten, die nicht für andere bestimmt seien und erzeuge dadurch für Dritte einen Mehrwert (Nassehi 2019, S. 302).

Noch schärfer ist da Soshana Zuboff: Die Verwendung von privaten Verhaltensdaten sei ein historischer Wendepunkt, der „Verhaltensüberschuss" wurde entdeckt und in ein marktfähiges Produkt und Geschäftsmodell verwandelt. „Die Nutzer sind die Quelle eines kostenlosen Rohstoffs für einen neuartigen Produktionsprozess" (Zuboff 2018/ 1 &2). „Unsere intimste alltägliche Realität ist im Überwachungskapitalismus wiedergeboren als Verhalten, das es zu überwachen und zu verändern, zu kaufen und zu verkaufen gilt" (Zuboff 2019, (3).

Nicht personenbezogene Daten

Nicht personenbezogene Daten können ihren Personenbezug durch Anonymisierung verloren haben. Ursprünglich personenbezogene Daten wurden so verändert, dass eine Zuordnung zu einer Person nur mit großem Aufwand noch möglich ist. Es ist also nicht sichergestellt, dass dieser Zustand durch Reanonymisierung wieder aufgehoben werden kann. Die Möglichkeit zur Identifizierung steigt logischerweise mit der Zahl der gespeicherten Datenmerkmale. Laut einer Studie in „Nature“, können mit nur 15 Merkmalen in jedem Datensatz, wie Alter, Wohnort oder Nationalität, 99,98 Prozent der US-Amerikaner identifiziert werden (https://netzpolitik.org/2019/).

Anonymisierte bzw. nicht personenbezogene Daten werden vor allem für Forschungszwecke genutzt, etwa für die Klimamodellierung, die lange Laufzeiten haben. Sie laufen auf Hochleistungsrechnern und erzeugen Datenmengen im Bereich Tera- und Petabyte für eine weltweite Nutzung, vorwiegend für wissenschaftliche und politische Debatten und u.a. auch für die Versicherungswirtschaft (Ludwig/ Thiemann 2020).

Kombination von personenbezogenen und nicht personenbezogenen Daten

Viele Anwendungen, wie das Internet der Dinge oder digitalisierte Autobahnen, liefern neben ihrer primären Verwendung durch Kombination personenbezogener und nicht personenbezogener oft Daten zur fortschreitenden Kommerzialisierung und Überwachung (Hesse 2020).

Hoffmann-Riem vermutet, dass mit Fortschreiten der digitalen Transformation die Kombination von unterschiedlichen Datensätzen stark zunehmen wird. Er veranschaulicht das anhand neuer Kfz-Modelle. Datensätze können erhoben werden vom Fahrverhalten des Nutzers und von seinem körperlichen Zustand, von den jeweiligen Straßenbedingungen, vom Zustand des Fahrzeugs etc. An diesen personen- wie nicht personenbezogenen Daten werden zahlreiche Akteure interessiert sein: Der Fahrer, um den Zustand des Wagens beurteilen zu können, der Anbieter des Navigationssystems wie staatliche Stellen, um aktuelle Daten über den Straßenzustand zu erhalten, Versicherungen, um Risiken und Unfallursachen bewerten zu können (Hoffmann-Riem 2022).

Die Kombination von personenbezogenen und nicht personenbezogenen Daten wird besonders für den Logistik- wie Produktionsbereich im globalen Wettbewerb überlebenswichtig sein. insbesondere im Bereich Industrie 4.0. Selbst wenn der überwiegende Teil als „reine Daten“ mit Sach- oder Materialbezug eingestuft wird, so ist in den Produktions- und Logistikketten häufig ein Anteil enthalten, der einen Arbeitsplatz oder menschliche Arbeitshandlungen einbezieht und der extrahiert werden kann.

Individuelle Interessen können darüber hinaus durch Verwertung

nicht personenbezogener Daten beeinträchtigt werden, indem diese mit Daten kombiniert werden, die dafür ursprünglich nicht gedacht waren, etwa mit anonymisierten Gesundheitsdaten oder individuellen Bewegungsdaten. Die Verfügungsmacht über diese Daten, der sich Internetnutzer kaum verweigern können und von der sie oft auch kaum etwas ahnen, generiert vor allem Macht bei den Plattformen über ökonomische und gesellschaftliche Prozesse und Entwicklungen, was tagtäglich an der Expansion von Google, Facebook & Co. erkennbar wird.

Intransparente Geschäftsmodelle

Naheliegend, dass sich mit entsprechenden Softwareangeboten und Beratungsleistungen viel Geld verdienen lässt. Für Unternehmen und Regierungen bietet beispielsweise das frühere Silicon Valley-Start-up *Palantir* – Investor u.a. der Deutsch-Amerikaner Peter Andreas Thiel - seine Software zur Analyse großer Mengen heterogener Daten erfolgreich an. Investoren oder Kunden sind u.a. der US-Geheimdienst CIA, Regierungen, das Militär und Behörden. Sie sorgen für rund die Hälfte des Umsatzes von 743 Millionen Dollar, was die Vermutung aufkommen lässt, dass staatliche Institutionen zum Zweck der Verschleierung ihrer Aktivitäten hier am Werke sind. Kunden in Deutschland sind der Pharmakonzern Merck, der Palantir-Software und –Beratung für molekularbiologische und pharmazeutische Forschungen einsetzt, das Medienhaus Springer, Polizeibehörden sowie Airbus, das darauf Wert legt, dass Palantir die Speicherung der Daten auf Servern in Europa zusichert (O.V. Handelsblatt Online 2020).

Schlussfolgerungen

Aus Daten gewonnene Erkenntnisse können sich positive wie negative Eingriffe in bestehende gesellschaftliche Strukturen ergeben. Negative, beispielsweise beim Schutz der Privatheit, wo es um Produktion vorhersagbarer Individuen durch Verhaltensüberwachung und –steuerung geht. Es können komplexe Strukturen in Ökonomie, Ökologie und Forschung transparent werden, die anders nicht offenzulegen sind. Die Mustererkennung in Verbindung mit Algorithmen der KI ist ein effektives Datenanalysewerkzeug, das sowohl zur Forcierung ökonomischer Produktivität, von Innovationssteigerungen und Machterhalt genutzt werden kann, genauso wie zur Zerstörung der Demokratie oder zur Vertiefung und Absicherung sozial-ökologischer Narrative.

Essay Arno Rolf

Literatur

Ulrich Dolata
Plattform-Regulierung. Koordination von Märkten und Kuratierung von Sozialität im Internet
Berliner Journal für Soziologie volume 29, pages 179–206 (2019)
Open Access Springer Link
https://doi.org/10.1007/s11609-020-00403-9

O.V. Handelsblatt Online (2020):
Überraschend mühselig und europäisch: Einblicke in das Geschäft von Palantir. Ein Insiderbericht
28.09.2020

Hesse, Wolfgang (2020):
Das Zerstörungspotenzial von Big Data und Künstlicher Intelligenz für die Demokratie, In: Informatik Spektrum(2020) 43:339-346 S.

Wolfgang Hoffmann-Riem:
Recht im Sog der digitalen Transformation
https://viewer.content-select.com/pdf/viewer?ip=91.54.176.10&id_type=isbn&identifiers=9783161612008&signature=9a396be27e9c96d981181eb952bfe4a4871e0fe9&frontend=1&language=deu
S. 32-35

Ludwig/ Thiemann (2020):
Datenkompetenz – Data Literacy.
Informatik Spektrum 45/2020

Nassehi Armin (2019):
„Muster. Theorie der Gesellschaft“, C. H. Beck,
München 2019.

Zuboff, Soshana (2019) (3)
http://www.faz.net/aktuell/feuilleton/debatten/die-digital-debatte/shoshana-zuboff-googles- ueberwachungskapitalismus-14101816.html?printPagedArticle=true#pageIndex_2

https://netzpolitik.org/2019/weitere-studie-belegt-luege-anonymer-daten/#vorschaltbanner, abgerufen 14.9.20).

Fragen:

(1) Das Unsichtbare sichtbar machen, wofür sollten wir es nutzen

(2) Welchen Sinn macht Datensensibilität?

(2) Beschreibt bitte die drei unterschiedlichen gesellschaftlichen Datentypen

Vertiefungsangebot:

Wolfgang Hoffmann-Riem.
Recht im Sog der digitalen Transformation
https://viewer.content-select.com/pdf/viewer?ip=91.54.176.10&id_type=isbn&identifiers=9783161612008&signature=9a396be27e9c96d981181eb952bfe4a4871e0fe9&frontend=1&language=deu

S. 32-35