

Sprechen Daten wirklich für sich selbst?

Ein Bericht vom BIG DATA-Symposium des Kulturinstituts an der Johannes Kepler Universität (JKU) in Linz, Österreich, 28./29. April 2017

Helga Bergmann

„Technische Innovationen wurden schon immer als Techno-Utopien und gleichzeitig als Sozial-Utopien verkauft“, stellte A.Univ.-Prof. Dr. Gerhard Fröhlich vom Institut für Philosophie und Wissenschaftstheorie der JKU in seiner Einführungsrede fest. Aussagen wie „Big Data enttarnt Terroristen unter den Flüchtlingen“, „mit Algorithmen gegen die Kriminalität“ oder „die heilende Kraft der Daten“ charakterisierten die PR-Botschaften, die aktuell in den Medien über Big Data zu finden seien. Eine kritische Auseinandersetzung mit dem Thema finde, so Fröhlich, viel zu selten statt. Mit dem Symposium „Big Data – Perspektiven kritischer Sozial- und Kulturwissenschaften“ wollten die Veranstalter deshalb „die Verheißungen der Big Data-Propheten auf den Prüfstand“ stellen. Dazu wurden die Methoden der Big-Data-Analytik beleuchtet und Auswirkungen hinterfragt, z.B. warum Bibliotheken in der EU-Perspektive zu Open Science gar nicht mehr erwähnt werden.



Der „Datengesellschaft“ wird von den Big Data-Protagonisten eine glorreiche Zukunft mit mehr Sicherheit, Bildung, Gleichheit und Emanzipation vorausgesagt. Zweifel daran meldete der wissenschaftliche Leiter des Symposiums, A.Univ.-Prof. Dr. Gerhard Fröhlich, in seinem Einführungsbeitrag „Die Verheißungen der Big Data-Propheten auf dem Prüfstand“ an. Big Data, ursprünglich

ein technischer Begriff für sehr große Datenmengen, werde heute vage und vieldeutig verwendet. Viele Big Data-Versprechungen seien nicht neu, betonte Fröhlich. Computer-Wissenschaftler hätten bei Einführung des Internets in den neunziger Jahren des vergangenen Jahrhunderts prophezeit, dass von nun an das Wissen der Menschheit jedem, der einen Computer hat, zur Verfügung ste-

hen und das Internet eine kollektive Intelligenz der Menschheit produzieren werde.¹ Empirische Studien zeichneten ein anderes Bild: Die Mehrzahl der das Internet nutzenden Menschen seien keine „prosumers“, sondern passive Nutzer/innen und selbst so kollaborative Projekte wie Wikipedia würden nur von wenigen, aktiven User/innen getragen. Auch der Zugang zum Wissen der Menschheit sei beschränkt, da sich fast alle relevante Daten unter privater oder staatlicher Kontrolle befinden würden und kaum oder nur stark zeitverzögert öffentlich zugänglich seien. Daher führe die „Datengesellschaft“ eher zu mehr Kontrolle und Machtkonzentration, konstatierte Fröhlich.

1 Zum Beleg für diese These siehe Fröhlich 1995, 1996 und die dort zitierten Quellen: Fröhlich, G. Demokratisierung durch Datenbanken und Computernetze? in: Thomas Becker et al. (Hg.), Informationsspezialisten zwischen Technik und gesellschaftlicher Verantwortung, S. 55-60, Stuttgart 1995, <http://hdl.handle.net/10760/8837>
Fröhlich, G. Netz-Euphorien. Zur Kritik digitaler und sozialer Netz(werk-)metaphern, in: Alfred Schramm, (Hg.), Philosophie in Österreich 1996, Wien 1996, 292-306, <http://sammel-punkt.philo.at:8080/1587/>

„Mit Facebook verkaufst Du Deine Seele“

In ihrem Vortrag „Wir kennen jeden Menschen besser als diese sich selbst“ untersuchte Magistra Silja Kempinger vom Kulturinstitut an der JKU die heute von fast allen Big Data-Dienstleistern benutzte Methode, mit der – gestützt auf Big Data-Auswertungen – Persönlichkeitsanalysen vorgenommen werden. Sie begann mit Cambridge Analytica (CA), einer amerikanischen Firma mit Sitz in London, die über große Bevölkerungsdatensätze verfügt und diese für Kunden aus Wirtschaft und Politik zielgerichtet auswertet und in Werbestrategien umsetzt.² CA rühmte sich u.a., am Erfolg von Donald Trump in den USA maßgeblich beteiligt gewesen zu sein. CA behauptet laut Kempinger, eine komplette Einflussnahme auf das Verhalten von Wählern sei möglich, wenn zusätzlich zu einer psychologischen Verhaltensanalyse von Menschen weitere Datenquellen wie Grundbucheinträge, Bonuskarten, Clubmitgliedschaften, Zeitschriftenabonnements oder sogar medizinische Informationen ausgewertet würden und dann Zielgruppen-gerichtete Werbebotschaften, sogenanntes Ad-Targeting, hinzukomme. Angeblich will CA Psychogramme von 220 Millionen Menschen erstellt haben. Mittlerweile wird angezweifelt, ob der Wahlausgang bei den US-Präsidentenwahlen auf diese Weise beeinflusst wurde, und auch CA selbst hat seine diesbezüglichen Aussagen relativiert. Kempinger erklärte, dass die Basis der von CA durchgeführten Persönlichkeitsanalysen das in den 1930er Jahren entwickelte OCEAN Modell ist. Dieses charakterisiert Menschen anhand von fünf

Eigenschaften: Offenheit (Openness), Gewissenhaftigkeit (Conscientiousness), Extraversion (Extraversion), Verträglichkeit (Agreeableness) und Neurotizismus (Neuroticism). Die heute auch von anderen Big-Data-Dienstleistungsunternehmen angebotenen Analyse-Methoden seien mittlerweile zwar durch technische Erweiterungen verfeinert worden, aber basierten immer noch auf dem OCEAN Modell, das wegen seiner groben Vereinfachung unter Wissenschaftlern umstritten ist.

Auf der Grundlage des OCEAN Modells wurde u.a. die Facebook-App MyPersonality entwickelt, die den Facebook-Nutzern als Spiel präsentiert wurde. Menschen aus rund 40 Ländern lieferten durch ihre Teilnahme an diesem Spiel zwischen 2007 und 2014 sechs Millionen Datensätze, die letztendlich als psychologische Verhaltensanalysen bei CA landeten. In einem nächsten Schritt wurden die Daten mit Informationen aus Facebook-Profilen und -Timelines verknüpft. CA behauptet, nach 68 Likes Hautfarbe, Intelligenz, Religionszugehörigkeit, Alkohol-, Zigaretten- und Drogenkonsum zu erkennen oder auch, ob die Eltern getrennt leben. Nach 150 Likes kenne das Datenunternehmen eine Person besser als deren Eltern sie kennen, und nach 300 besser als die Person sich selbst.

Die Referentin rief dazu auf, mehr Eigenverantwortung bei der Nutzung des Internets und seiner Dienste zu zeigen: „zuerst denken und dann klicken“. Besondere Vorsicht sei bei Facebook geboten, da es tatsächlich alles registriere. Auf keinen Fall sollten Facebook-Handy-Apps verwendet werden, da sie reine Spyware-Werkzeuge seien. Als Beispiel für die Skrupellosigkeit von Facebook bei der Verwendung von privaten Daten verwies Kempinger auf das Psycho-Expe-

riment 2012, das unter wissenschaftlicher Aufsicht, aber ohne Wissen der Teilnehmer durchgeführt wurde. Facebook hatte damals fast 700.000 Menschen zu Versuchskaninchen gemacht, indem sie der Gruppe 1 vorwiegend negative News und der Gruppe 2 hauptsächlich positive News zeigten. Gruppe 1 postete daraufhin mehr negative, Gruppe 2 mehr positive Nachrichten. Das Ergebnis des Experiments wurde in den Proceedings of the National Aca-



Magistra
Silja Kempinger

demy of Science³ veröffentlicht. Kempinger beendete ihren Vortrag mit einem Hinweis auf die Daten-Besitzverhältnisse: „Den Zugriff auf die Daten haben nur wenige und die analysieren uns, nicht wir sie!“

Nicht viele, sondern die richtigen Daten zu haben, ist wichtig

In ihrem Vortrag „David gegen Goliath – oder welche Datenbasis bringt welche Erkenntnisse“ zog Magistera Dr.in Andrea Ghoneim einen Vergleich zwischen quantitativen und qualitativen Forschungsansätzen. Die wissenschaftliche Mitarbeiterin am Department für

2 “Data drives all that we do. Cambridge Analytica uses data to change audience behavior.” <https://cambridgeanalytica.org/>

3 Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci USA 2014; 111 (24): 8788-8790. doi: 10.1073/pnas.1320040111

Interaktive Medien und Bildungstechnologien an der Donau-Universität Krems, Österreich, wies darauf hin, dass quantitative Darstellungen den qualitativen Darstellungen den Rang abzulaufen scheinen, was sie bedauerte. Zu welcher unterschiedlichen Aussagen die beiden Forschungsansätze kommen, erläuterte Ghoneim anhand des EU-Projekt ATS2020, an dem sie beteiligt ist. ATS2020 evaluiert ein neues Lernmodell im Hinblick auf die Entwicklung fächerübergreifender Fähigkeiten bei Schülern und Schülerinnen im Alter von 10 bis 15 Jahren. Insgesamt nehmen derzeit 8.600 Kinder aus zehn europäischen Ländern daran teil. Sie sind zwei Gruppen, einer experimentellen und einer Kontroll-Gruppe, zugeteilt. Nur die experimentelle Gruppe wird nach dem neuen Lernmodell unterrichtet. Das Projekt verfolgt einen multimethodischen Ansatz mit Desktop Recherche (Literaturarbeit), qualitativer und quantitativer Forschung. Die quantitative Auswertung erfolgt anhand von Tests vor und nach der Einführung des neuen Lernmodells. Dieses Ergebnis zeigt in Prozent, ob die Schülerinnen und Schüler von dem neuen Unterrichtsmodell profitieren. Weit aufwändiger ist die qualitative Forschung. Hier werden durch Beobachtungsprotokolle, Fragebögen für Lernende und Lehrende, Selbst-Evaluierungen, Bewertung von Artefakten und Leitfaden-Interviews inhaltliche Fragen beantwortet: Hat das neue Lernmodell fächerübergreifende Fähigkeiten vermittelt und mit welchem Erfolg? Die Daten, die in diesem Großexperiment gewonnen und verarbeitet werden, liegen dennoch in einem überschaubaren Bereich, betonte Ghoneim. Anders sei das bei Learning Analytics zur Selbst-Evaluierung. Die Daten werden ähnlich wie bei

Twitter oder Facebook über eine Schnittstelle (API/Application-Programming Interface) aus den Lernprogrammen gezogen. Es entstehen Datenmengen in einer Größenordnung, die nur noch von Computernetzwerken bearbeitet werden kann. Bei einem Test mit der Open Source Lern- und Selbst-evaluierungs-Software xApi innerhalb des Moodle der Donau-Universität Krems wusste aber am Ende niemand, so Ghoneim, was man mit der Menge an gewonnenen Daten genau machen könnte.

In ihrer Einschätzung von Big Data bezog sich Ghoneim auf Lev Manovich⁴, der davon ausgeht, dass eine Vertiefung der quantitativen Forschung durch Big Data denkbar sei, indem man die Datenströme verfolgt und sie zurückrechnet. Auch bestünde die Möglichkeit, dass sich aus Big Data ein neues Forschungsparadigma entwickelt. Zum gegenwärtigen Zeitpunkt sei dies jedoch noch Spekulation. Solange die Kulturwissenschaft die Fragen und die Informatik die Daten habe, bestehe ein „digital divide“. Es sei daher wichtig, dass Forschende in den Sozial- und Kulturwissenschaften die Auswertungsmöglichkeiten von Big Data kennen. Hilfe dabei bietet beispielsweise das Open Source Programm „R“. In der Einleitung zum Buch „R for Data Science“⁵ schreiben die Autoren: „Wenn Eure Daten größer sind, dann überlegt euch sorgfältig, ob euer Big Data Problem vielleicht nur ein verkleidetes Small Data Problem ist. Während die kompletten Daten groß sind, sind vielleicht die Daten, um eine Frage

zu beantworten, ohnehin nur kleine Daten.“ Und damit, so die Referentin haben die Autoren Klartext gesprochen: Es ginge letztendlich nur darum, die richtigen Daten zu haben und nicht darum, wie viele man hat. Und auch dazu haben die Autoren die Antwort: „Die Herausforderung besteht darin, die richtigen Small Data zu finden und das erfordert oftmals eine Vielzahl von Iterationen.“ Ihre persönliche Überzeugung zu Big Data brachte Ghoneim mit der Frage zum Ausdruck: Wenn Big Data die Antwort ist, was ist dann die Frage? Zum Schluss ihres Vortrags forderte Ghoneim die Geisteswissenschaftlerinnen und Geisteswissenschaftler auf, kritisch Stellung zu den Problemen mit Big Data zu nehmen, auch wenn das gegenwärtig eher anmutet wie der Kampf David gegen Goliath.

„So etwas wie ROHdaten gibt es nicht“

Professor Dr. Gerhard Fröhlich machte in seinem Vortrag „Rohes & Gekochtes informationswissenschaftlich“ einen Ausflug ins Kulinarische. Nicht nur Speisen können roh, halbgar oder gekocht zubereitet werden, sondern auch Daten, Konzepte und Texte. Sprechen Wissenschaftler etwa von Rohdaten, Rohkonzepten oder Rohtexten, so ist das, laut Fröhlich, eine Ausrede dafür, dass etwas noch nicht fertig, noch nicht zitierfähig ist, eine Art Immunisierung gegenüber Kritik. Ein halbbares Konzept ist unausgereift, noch nicht zu Ende gedacht. Gekocht würden Daten allerdings schon bei ihrer Erhebung, weswegen Daten keine Naturressourcen sein könnten, wie immer öfter zu lesen sei (vgl. die Parole „Daten als Rohstoffe“). Einerseits sind in Forschungsprozessen laufend Entscheidungen zu fällen, die die Daten beeinflussen, insofern würden Daten genauso

4 Manovich L. Trending - Verheißungen und Herausforderungen der Big Social Data in: Reichert R. (Hrsg.): Big Data: Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie. transcript Verlag, Bielefeld 2014, S. 65 ff. ISBN 978-3-8376-2592-9

5 Wickham H, Grolemund G. R for Data Science. Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media 2017. 520 Seiten



„gekocht“ wie eine Mahlzeit in der Küche (Was betrachte ich als auf Daten reduzierbar, was erkenne ich als Datum an, welche „Ausreißer“ entferne ich usw.).

Zugleich steht der englische Begriff „cooking“ data, seit 1830 der britische Mathematiker Charles Babbage diesen Begriff einführte, auch für Daten fälschen, entweder durch Zurückhalten von Daten oder die Veröffentlichung nur jener Daten, die der Intention entsprechen. Heute sei diese Form der Manipulation von Daten durchaus gängige Praxis. Diese Entwicklung sei nicht zuletzt dem „Impact-Wahn“ zuzuschreiben, betonte Fröhlich, der ein Verfechter obligatorischer Forschungsregister ist, wie sie in der Medizin in den USA auch bereits eingeführt seien: Müssten bereits im Vorhinein die Hypothesen, Art und Größe der Stichproben, Zeitraum der Erhebung, statistische Auswertungsverfahren genannt werden, könnten Forschungsgruppen nicht hinterher Hypothesen einfach umdrehen, Stichproben unter den Tisch fallen lassen oder den Zeitpunkt des Erhebungsendes manipulieren. Die permanente Evaluie-

rung der Wissenschaft, der Zwang als Erfolgsbeweis möglichst viel zu publizieren, veränderten wissenschaftliches Arbeiten und Forschen. Zudem publizierten viele TOP-Journale nur hoch korrelative, signifikante und sensationelle Ergebnisse auf der Jagd nach möglichst hohen Impact Faktoren (= in den letzten zwei Jahren durchschnittlich erwirtschaftete Zitationen). Studien, in denen eine Hypothese nicht verifiziert wurde, würden kaum gedruckt. Doch es gibt Hoffnung: Einige innovative Journale referieren die Studien noch vor Bekanntwerden der Resultate, und versprechen der Forschungsgruppe, die Studie in ihrem Journal zu veröffentlichen, ganz egal ob sie positive, negative oder Nullresultate erbringen würde.

Bei Big Data sei davon auszugehen, dass nur ein Bruchteil der Daten überhaupt verwendungsfähig ist. Er verwies dabei unter anderem auf die Doktorarbeit von Terje Tüür-Fröhlich⁶, die die Fehler-

⁶ Tüür-Fröhlich T. The Non-trivial Effects of Trivial Errors in Scientific Communication and Evaluation. Schriften zur Informationswissenschaft 69. Verlag Werner Hülsbusch, Glückstadt 2016. ISBN 978-3-86488-104-6

rate bei der Zitationsdatenbank SSCI (Social Science Citation Index) untersucht hat. In einem Beitrag der renommierten Zeitschrift Harvard Law Review lag die Fehlerrate bei den im Original korrekten bei 99 Prozent. Auch beim Ngram-Viewer von Google Books deuten erste kritische Untersuchungen eine geringe Rate an brauchbaren Daten an, in einem Fall, je nach Fehlertoleranz, von 8 bis 2%.⁷ Beide Systeme (WoS wie Google Books) hätten massive SCAN-, OCR- und Parserprobleme.

„Werkzeugkästen oder Durcheinander?“

A.Univ.-Prof. Dr. Herbert Hrachovec vom Institut für Philosophie der Universität Wien versah seinen Vortrag zur Datenverarbeitung mit dem Titel „Schubladen und Wolkenfenster“. Bevor Hrachovec auf Big Data und die sich daraus ergebenden möglicherweise neuen Aspekte zu sprechen kam, warf er einen Blick auf die

⁷ Rosenberg, D. Data before the fact., in: Gitelman L (ed.) „Raw Data“ is an Oxymoron. The MIT Press Cambridge, Mass./London, England 2013, 29

Datenverarbeitung in Datenbanken. Über 90 Prozent der Datenverarbeitung, die im Alltag eine Rolle spielt, basiert, so der Referent, auf dem Grundschemata einer relationalen Datenbank. Dabei ist das relationale Datenmodell eine geordnete, in einer Tabellenform darstellbare Beziehung zwischen verschiedenen Faktoren. Philosophisch betrachtet sind die Reihen in einer Tabelle Kategorien und die Rubriken Schubladen, um bei Hrachovec's Bild zu bleiben. Jede



Dr. Gerhard Fröhlich

Zeile hat eine intendierte Bedeutung und kann als Prädikat (= Satz-aussage) verstanden werden, das, in einem Satz formuliert, wahr oder falsch sein kann. Datenbanken kann man, so Hrachovec, als Ansammlung von Sätzen verstehen. Dabei bildeten die Sätze modellhaft die Welt ab.

In einer Art Manifest zu Big Data hat sich Chris Anderson⁸ 2008 gegen Modelle und für Big Data zur Erklärung der Welt ausgesprochen. Die weltweite Datenerzeugung (auf Servern von Google und NSA!, Anm. des Referenten) liefert eine Sintflut an Daten, die sich ständig veränderten und sich da-

her der Modellbildung entzogen. Deswegen propagiert Anderson, sich von der Datenfülle inspirieren zu lassen und mithilfe von Algorithmen mögliche Zusammenhänge herauszufinden und so zu Erkenntnissen zu kommen.

Für Hrachovec stellt sich hier die Frage, ob die Daten im Netz als eine Art Rohdaten einen direkteren Zusammenhang mit dem, was die Welt ist, erlauben, als die vorstrukturierte Herangehensweise. Der Vorteil könnte sein, dass die Auswertung ohne Vorstruktur, sprich ohne Vorurteil, erfolgt. Dass dem aber nicht so ist, erklärte Hrachovec am Beispiel eines Werkzeugkastens. Dabei ist der klassische Datenbankansatz mit den Schubladen der geordnete Werkzeugkasten, wohingegen die Datenwolke (Big Data) eine Kiste mit einem Durcheinander von Werkzeugen sei, von der man nicht wisse, was sich darin verbirgt. Entweder man findet mit „tryanderror“ heraus, welchen Inhalt die Kiste hat, oder man macht eine Mustererkennung mittels Algorithmen und einer latenten semantischen Analyse. Dabei könne man zu den unterschiedlichsten Ergebnissen kommen: einmal, dass es sich um Werkzeuge handelt; genauso gut kann herauskommen, dass es eine Anzahl roter, blauer und gelber Gegenstände ist oder ein Test zur Schwerelosigkeit, der mit den in der Kiste befindlichen Werkzeugen durchgeführt wird. Wenn man etwas herausfinden will, das von Interesse ist, dann muss man Begriffe, also Schubladen haben, ist Hrachovec überzeugt.

Wissenschaftler und Wissenschaftlerinnen als Big Data-Objekte

Mit dem Thema „Metriken, Kennziffern, Event Logging – Informationswissenschaft pro und contra Wissenschaftssteuerung“ setzte

sich Dipl.-Soz. Dr. Ulrich Herb von der Universitätsbibliothek Saarbrücken auseinander. Big Data, so der Referent, stelle neue Anforderungen an die Qualifikation der Wissenschaftlerinnen und Wissenschaftler, ablesbar an der großen Nachfrage nach Data Scientists. Wer mit Big Data arbeitet und die Daten nutze, handle selbst, sei Subjekt. Big Data würde Wissenschaftlerinnen und Wissenschaftler aber auch zu Objekten zu machen. Dies geschehe über Protokollierung, Kategorisierung, Belohnung und Prognosestellung. Mittlerweile wird, so Herb, auch der Wert einer Universität anhand bibliometrischer Daten, eingeworbener Drittmittel und Anzahl der Promotionen gemessen. Neu sei in diesem Zusammenhang das sog. Event logging, das zu einem lohnenden Geschäftsfeld geworden sei. Dabei handelt es sich um eine Methode, die digitalen Spuren eines Wissenschaftlers, einer Wissenschaftlerin zu protokollieren und dabei auch die Rezeption durch andere zu berücksichtigen, seien es die wissenschaftliche Community, Medien oder die Öffentlichkeit. Das Geschäftsfeld von Metriken, Kennziffern und Event Logging werde nicht nur, aber prominent von Elsevier angeboten. Der Wissenschaftsverlag habe sich zu einem „Global Information Analytics“-Unternehmen entwickelt. Mit Science Direct hat Elsevier schon vor zwei Jahrzehnten mit dem Aufbau einer Volltext-Datenbank für begutachtete wissenschaftliche Literatur und mit Scopus einer ebensolchen Abstracts- und Zitations-Datenbank begonnen. 2009 folgte dann SciVal, ein Werkzeug zum Benchmarking von Wissenschaftseinrichtungen. 2012 wurde Pure eingeführt, ein Instrument, das es ermöglicht, die Wissenschaftlerinnen bzw. Wissenschaftler innerhalb einer

⁸ Anderson C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. <https://www.wired.com/2008/06/pb-theory>

wissenschaftlichen Einrichtung zu vergleichen. 2013 erfolgte der Kauf von Mendeley, einem System zur Verwaltung wissenschaftlicher Publikationen, das Elsevier zusätzlich mit Informationen füttert, z.B.: In welchen Wissenschaftsbereichen werden Versuche durchgeführt? Wie lange wurde ein pdf gelesen? etc. 2016 erwarb Elsevier das Repository SSRN, in dem man Arbeitspapiere und Vorabpublikationen in den Geisteswissenschaften deponieren kann. Das Unternehmen hat auf diese Weise Einblick in die Forschungstätigkeiten und Publikationsgepflogenheiten. Im selben Jahr tätigte Elsevier mit Hivebench vielleicht einen der wertvollsten Einkäufe, mutmaßte Herb, denn damit verwalten Naturwissenschaftlerinnen und Naturwissenschaftler ihre Experimente, Protokolle und Analysen. In einem nächsten Schritt kaufte Elsevier im Februar 2017 Plum Analytics, einen Pionier auf dem Gebiet von Altmetrics. Damit wird die Rezeption einer wissenschaftlichen Publikation, aber auch von wissenschaftlichen Einrichtungen sowie Wissenschaftlerinnen und Wissenschaftlern zusätzlich zum wissenschaftlichen Kontext auch in den sozialen und Publikumsmedien gemessen.

Wenn die Informations-Wissenschaft, aber auch die Wissenschaft insgesamt, mit Altmetrics und ähnlichen Messverfahren, die methodisch durchaus angreifbar seien, bewertet und geleitet wird, besteht, so Herb, die Gefahr der Fremdsteuerung der Wissenschaft und einer dirigistischen Forschungsplanung. Denn auch die Mittelvergabe orientiert sich an den Erkenntnissen aus Metriken, Kennziffern und Event Logging. So schreibt beispielsweise der Forschungsförderungsfond zwingend die Verwendung der Datenbanken von Elsevier, Springer

und ähnlichen Anbietern vor. Vor diesem Hintergrund sei es nicht verwunderlich, dass Bibliotheken in der EU-Perspektive 2016 „Open innovation, open science, open to the world – a vision for Europe“, gar nicht mehr erwähnt würden.

Big Data und logic driven innovation: das größte logische Experiment auf Grundlage von 0 und 1

Einen zugegebenermaßen düsteren Blick in die Zukunft warf Privatdozent für Operations Research & Knowledge Management (St. Gallen) Dr. Dirk Solte in seinem Vortrag „Big Data-Analytik – Auswirkungen auf Beschäftigung und Teilhabe“. Solte ist Leiter des Verlags der Versicherungswirtschaft Karlsruhe, und war vorher langjähriger Forscher am Forschungsinstitut für angewandte Wissensverarbeitung in Ulm, kennt die „Künstliche Intelligenz“ also als Entwickler.

Um zu zeigen, wo Big Data heute steht, benannte Solte die Ebenen der Informationsverarbeitung, die mit den vier Ebenen der menschlichen Kognition korrespondierten: Die unterste Ebene ist die der direkten Interaktion mit der Umwelt, die oberste, vierte, die Ebene der Theorie. Die beiden subsymbolischen Ebenen sind die neuronale und die Signal-Ebene. Die beiden symbolischen Ebenen beinhalten Informationsverarbeitung und Regelverarbeitung. Laut Solte ist Big Data jetzt an einem Punkt angelangt, an dem begonnen wird, auch die vierte Ebene, die der menschlichen Kognition entspricht, zu automatisieren, d.h. alle vier Ebenen können maschinell ausgewertet werden. Bisher hatte die Automatisierung geringer qualifizierte Berufe betroffen, nun aber geht es um Rechtsanwälte, Rechtsanwältinnen, Ärzte und Ärztinnen und andere hochqualifizierte Berufe. Die Erfolge neuerer GO-Program-

me gegen die menschlichen GO-Meister würden die Leistungsfähigkeit dieser selbstlernenden Maschinen zeigen.

Das habe aber Auswirkungen auf Teilhabe, Beschäftigung und die gesellschaftliche und globale Stabilität, vor denen wir unsere Augen nicht verschließen dürften.

Heute leben 7,5 Milliarden Menschen auf der Erde. Die Vereinten Nationen (UNO) rechnen pro Jahr mit einer Zunahme der Weltbevölkerung um 78 Millionen. Das



PD Dr. Dirk Solte

Versprechen auf zukünftige Wertschöpfung wachse schneller, als es die Menschheit einlösen könne, konstatierte Solte. Bereits heute fordere die Menschheit der Natur pro Jahr 1,5 Jahre ihrer Leistungskraft ab. Eine Verzehnfachung der Wertschöpfung in den nächsten 20 Jahren, wie vorausgesagt, ist daher laut Solte unmöglich und selbst wenn sie möglich wäre, würde sie nicht ausreichen, um die durch den nächsten Innovationschub wegfallenden Arbeitsplätze durch neue zu ersetzen. Von den durch die maschinellen Analysemöglichkeiten wegfallenden Arbeitsplätzen sei vor allem

die Mittelschicht betroffen, die in der bisherigen Gesellschaft der Garant für Stabilität war. Zur weltweiten Instabilität trage bei, dass derzeit 1,35 Milliarden Menschen mehr als 80 Prozent der Weltjahresproduktivität beanspruchten und 6,2 Milliarden Menschen sich weniger als 20 Prozent der Weltjahresproduktivität teilen müssten. Vor diesem Hintergrund sei auch die Flüchtlingswelle zu erklären. Gelingt es nicht, verbindliche Standards für Umwelt und Soziales bei der Produktion von Waren und Dienstleistungen und die autonome Hoheit der Daten auf globaler Ebene einzuhalten, befürchtet Solte einen Kollaps.

Neue Bilder aus dem All – Visualisierung von Daten & Modellen, willkürlich eingefärbt

Magister Andreas Vogl, Dissertant und Lektor am Institut für Philosophie und Wissenschaftstheorie an der Linzer JKU, zeigte in seinem Beitrag „Datenbilder und der Kampf um Aufmerksamkeit“ schöne Bilder aus der Astronomie. Hinter diesen Bildern, deren Botschaft nach Möglichkeit auch noch publikumsträchtig sein sollte, stehe der Wettbewerb um Fördermittel. Dies gelte in besonderem Maß für die Astronomie, die zumindest derzeit weder militärisch noch ökonomisch sonderlich interessant sei und daher immer nachweisen müsse, dass ihre teuren Teleskope, Forschungssatelliten und Radarstationen und die damit produzierten Untersuchungsergebnis-

se auch gesellschaftlich relevant sind.

Auf dem Weg in die Publikumsmedien machen die astronomischen Bilder, die ursprünglich der Kommunikation zwischen Experten dienten, einen Transformationsprozess durch, erklärte Vogl. Sie würden zu einem PR-Projekt, für das andere Anforderungen bestünden als für die wissenschaftliche Abbildung: Das Bild müsse ästhetisch und ein Eyecatcher sein. Ein weiteres Problem bei wissenschaftlichen Bildern sei, dass ihnen häufig eine sehr große Datenmenge zugrundeliegt. Nur durch Bearbeitung, Reduktion und Vereinfachung werden sie als Bild darstellbar, was den Eindruck vermittelt, es handle sich um eine Fotografie. Als Beispiel nannte Vogl die von der Mars-Sonde aufgenommenen Panorama-Bilddaten des Planeten, die so zusammengeschnitten und bearbeitet wurden, dass sie als „Schnappschüsse“ vom Mars durch Fernsehen, Zeitschriften und Zeitungen gingen. Um die Bilder vom Mars eindrucksvoll zu gestalten, wurden Farbanpassungen vorgenommen: Über dem rostbraunen Mars

wurde mittels Weißabgleich eine Oberfläche erzeugt, die aussieht, als gäbe es dort eine Atmosphäre. Mittels Blaufärbung werde ein blauer Himmel suggeriert, so dass die Assoziation entsteht, es gäbe dort Wasser. Die Vermutung liegt nahe: Hier soll unterschwellig für eine künftige Kolonisierung des Mars geworben werden. Wenn der Entstehungsprozess solcher Bilder nicht bekannt gemacht wird, also nicht klar wird, dass es sich um visualisierte Daten handelt, werden Bilder als Fotografien wahrgenommen. Und Fotografien hätten selbst im Zeitalter von Digitalkameras immer noch den Nimbus der Objektivität und Authentizität. Am Ende des Symposiums blieb der Eindruck, hochspannende Informationen und berechtigte Kritik gehört zu haben. Aber wie sollen die praktischen Konsequenzen aussehen? Um Risiken und Chancen von Big Data zu beleuchten, bräuchte es auch Begegnungen zwischen den „Machern“ und den „Kritikern“. Nur wenn alle Disziplinen zusammenarbeiten, wird es möglich sein, das Potenzial von Big Data zu nutzen und seine Auswüchse abzuwenden. **I**



Helga Bergmann-Ostermann

Medizinjournalistin

Dipl.-Übersetzerin

h.bergmann-ostermann@t-online.de