

Klaus Wiegerling

Big Data in der Digitalen Informationsgesellschaft (IKGIS-Workshop, 7.3.2018)

Der Kurzvortrag versucht einen Einblick zu geben in einige Grundprobleme bei der Anwendung von Big-Data-Technologien wie etwa dem maschinellen Lernen. Die Absicht große fließende Datenmassen zu strukturieren und aus ihnen Erkenntnisse zu gewinnen ist mit vielen Problemen verbunden, die Auswirkungen sowohl auf den wissenschaftlichen Gebrauch als auch auf den Gebrauch der Technologie im praktischen Alltag haben. Auffällig im gegenwärtigen Big-Data-Diskurs sind Fehleinschätzungen, die aus begrifflichen Unschärfen resultieren. So ist die Rede von Daten als Rohstoff ebenso problematisch wie die Annahme, dass man von der Wirklichkeit ein Datendouble erzeugen könne. Die Verwechslung von Wirklichkeit und Modell ist weit verbreitet. Auch die ungenaue Unterscheidung von Datum, Information und Wissen fördert viele Fehleinschätzungen und Missverständnisse. Zuletzt soll gezeigt werden, dass die einst von der Wissensgesellschaft überwunden geglaubte Informationsgesellschaft derzeit eine Unterbietung durch die Datengesellschaft erfährt. Von entscheidender Bedeutung für die Nutzung und ethische Beherrschung von Big-Data-Technologien sind deren Einbettung und unsere Positionierung gegenüber der Technologie und ihren Ergebnisangeboten.

Meine Damen und Herrn,

Im aktuellen Diskurs gilt Kritik an der Digitalisierung und Vernetzung als hinterwäldlerisch. Ich erinnere mich als ich vor Jahren eine Schockstarre bei einem ORF-Gespräch bewirkte als ich das Wort Entnetzung gebrauchte. Wir reden von Autofasten, vom Informations- und Kommunikationsfasten reden wir seltsamerweise nicht, obwohl bekannt ist, dass - wie es in der FAZ hieß – das Netz einen dicken Auspuff hat.

Ich will nun entsprechend der Kernaufgaben der Philosophie: Kritik, Vermittlung und Orientierung einige Bemerkungen zu aktuellen informatischen Entwicklungen machen, in denen ich eine kritische Bewertung der Anwendungsmöglichkeiten vornehmen will. Weiterhin will ich Bewertungskriterien nennen, die auch bei der Nutzung von Daten und Informationssystemen in den Zusammenhängen, die bei dieser Tagung behandelt werden, von Bedeutung sein können. Ich werde dabei den Rahmen des Tagungsthemas übersteigen, aber damit auch eine kritische Einbettung der Digitalisierungsdebatte vornehmen, die für den konkreten Umgang mit fortgeschrittenen Informationssystemen produktiv sein kann.

Kritik artikuliert sich in einer begrifflichen Unterscheidungspraxis, die in der gegenwärtigen Debatte oft vernachlässigt wird.

Als Schlüsselproblem im Umgang mit komplexen Datenmassen und Systemen, die diese strukturieren sollen, sehe ich, dass Entlastung und Gewährung neuer Handlungsoptionen, durch die Technik gerechtfertigt wird, zunehmend in eine Entmündigung und Einschränkung von Handlungsmöglichkeiten umschlagen, zumal in vielen Bereichen der Ausschluss des Menschen aus Produktions- und Organisationsabläufen als Ausdruck besserer Funktionalität und höherer Objektivität gilt.

Überlastung ist ein chronischer Zustand komplexer datengetriebener Gesellschaften, in denen ständig neue Datenberge aufgehäuft werden. Mit Hilfe unterschiedlicher informatischer Werkzeuge versucht man die bestehende Unübersichtlichkeit zu strukturieren und Komplexität zu reduzieren. Es geht bei der Anwendung von Big-Data-Technologien sowohl um die Gewinnung von Erkenntnis, als auch um die praktische Entlastung unseres Alltagslebens, in dem Entscheidungen getroffen und Verrichtungen in begrenzter Zeit erledigt werden müssen.

Selbstbestimmte Entscheidungen setzen Wahlmöglichkeiten voraus, ein zu viel an Wahlmöglichkeiten kann Wahl aber auch verunmöglichen. Es gibt keine Entscheidung ohne, aber auch keine in einem Meer unüberschaubarer Wahlmöglichkeiten. Diese Einsicht hat nicht nur eine psychologische Dimension, sondern auch eine ethische. Entscheiden im Sinne einer bewussten Wahl setzt überschaubare Alternativen voraus. Um in einer komplexen Gesellschaft bestehen zu können sind wir auf komplexitätsreduzierende Systeme angewiesen.

Im Falle spezieller Big-Data-Technologien werden Ergebnisangebote gemacht, deren Zustandekommen wir nicht oder nur mit hohem Aufwand rekonstruieren können. Die für wissenschaftliche Geltungsansprüche

essentielle Rekonstruierbarkeit der Ergebnisse ist infrage gestellt, wenn sie mithilfe selbstlernender Algorithmen gewonnen werden. Das Ergebnis kann zwar durch eine andere Datenanalyse bestätigt werden, die Gründe des Zustandekommens können aber unterschiedlich sein. Sybille Anderl nennt in ihrem Artikel „Künstliche Intelligenz: Denn wir wissen nicht, wie sie’s tun“ einen Fall, bei dem es um die Identifizierung von Pferdebildern ging: „So untersuchten (...) Wissenschaftler (...) die Ergebnisse zweier Algorithmen zur Bildklassifikation, die (...) trotz verschiedener (...) Methoden (...) eine ähnliche Genauigkeit erreichten: Der erste Algorithmus beruhte auf einem tiefen neuronalen Netzwerk, während der zweite eine andere Methode des Maschinenslernens nutzte. Als die Forscher sich die Hintergründe der Klassifikation genauer ansahen, kam heraus, dass beide Algorithmen verschiedene Strategien entwickelt hatten. Während der Netzwerk-Algorithmus sich (...) an Konturen im Bild orientierte (...), orientierte der andere sich an einer kleinen Schrift, die ausschließlich in Pferdebildern auftauchte und dort das Copyright der Bilder angab. Der zweite Algorithmus hatte zwar richtig erkannt, dass sich im Testsample Pferdebilder durch dieses Copyright-Zeichen auszeichneten. Dass diese Eigenschaft (...) nichts mit Pferden zu tun hatte, konnte der Algorithmus nicht bemerken. Bei einer Anwendung auf beliebige Pferdebilder hätten beide Algorithmen daher (...) sehr unterschiedlich abgeschnitten.“

Das Beispiel zeigt, dass der Einsatz selbstlernender Systeme für die wissenschaftliche Erkenntnisgewinnung problematisch sein kann. Ein an sich richtiges Ergebnis kann aus falschen Gründen zustande kommen. Ist es deshalb falsch? Nein, es ist aber wissenschaftlich gesehen auch nicht richtig. Wissenschaftlichen Ansprüchen kann nur genügen, was methodisch gewonnen wurde und auf gleiche oder ähnlich gelagerte Fragestellungen übertragen werden kann. Es kann nur Geltung beanspruchen, wenn es auf einer Erfassung der Fragestellung beruht. Die Nachvollziehbarkeit eines Ergebnisses ist aber im Falle der Nutzung

selbstlernender Systeme nicht mehr gegeben. Sie verändern permanent ihren Wissensstand ,und‘ das Verfahren um zu Ergebnissen zu erlangen. Eine exakte Wiederholung des Ergebnisses ist nicht gewährleistet.

Systeme legen Ergebnisse vor, die als Entscheidungen Anerkennung finden. Im strengen Sinne fällen sie jedoch keine Entscheidung, da diese von ihnen nicht verantwortet wird und für sie keine Folgen hat. Wir können aber in einem weiteren Sinne von einer Verantwortlichkeit des Systems reden, wenn wir davon ausgehen, dass es ein repräsentativer Ausdruck der Gesellschaft ist, die sich in kulturellen und sozialen Präferenzen artikuliert. Im Sinne von Bruno Latour ließe sich von einer Systemverantwortung reden, insofern er technische Systeme als Netzwerke von gesellschaftlichen, personalen und technischen Elementen sieht, die wie der Mensch unter dem Aspekt des Handelns gesehen werden können. Latours Überlegungen weisen auf die Gesellschaft und Individuum disponierende Rolle technischer Systeme. Sogar in einem historischen Sinne wird Technik damit zur Vermittlungsinstanz. Es stellt sich aber die Frage, inwiefern durch technische Systeme Werte gesetzt und Positionen eingenommen werden. Als Verkörperungen gesellschaftlicher und gruppenspezifischer Wertsetzungen motivieren sie zwar bestimmte Wertnahmen, bringen aber selbst keine Werte hervor, solange sie keine eigenen Intentionen verfolgen.

Unter dem Schlagwort Big Data artikuliert sich eine Massenideologie, die die Entwicklung der Informationstechnologien von ihren Anfängen an in unterschiedlicher Stärke begleitet. Die Idee der Datafizierung, dass jede physische, gesellschaftliche und psychische Äußerung in Daten erfassbar sei und ein mehr an Daten zu quasi ,wahreren‘ Erkenntnis führe, ist freilich nur bedingt richtig. Bei jeder Datenanalyse finden Artikulationen und Desartikulationen von Daten statt. So richtig es ist, dass die permanente Aktualisierung und Vermehrung von Messdaten Einfluss auf die Qualität

eines Ergebnisses haben kann, so falsch ist es, dass ein mehr an Daten generell zu besseren Ergebnissen führe. Erkenntnis kommt nicht durch ein Datendouble zustande, sondern durch die Betonung von als wesentlich erachteter Momente und die präzise Bestimmung der Perspektive.

Einer der verbreitetsten Glaubenssätze des Big-Data-Diskurses ist, dass man mit Hilfe von Datenmassen das Singuläre bzw. Ereignishaftes erfassen und damit die Notwendigkeit eines notorisch unscharfen hermeneutischen Zugangs überwinden könne. Die klassische Erkenntnis ‚De singularibus non esse scientia‘ soll mit informatischen Verfahren widerlegt werden. Man differenziert aber mit mehr Daten nur eine Typologie aus. In personalen Kontexten kann es gefährlich werden, wenn man die Differenz zum Singulären und Ereignishaften nicht mehr erkennt und Personalität auf einen Datensatz und der darin enthaltenen Relationen begrenzt, ihren transzendierenden und wandelbaren Charakter aber verkennt.

Damit will ich einige Kriterien nennen, die m.E. an die Diskussion über den verantwortungsvollen Umgang mit Big-Data-Technologien anzulegen sind.

Zunächst will ich auf die Unterscheidung von rechtlichem und ethischem Diskurs eingehen, da es hier immer wieder zu Missverständnisse kommt, was daran liegt, dass Ethikdebatten zunehmend auf Konsensfindung reduziert werden. Das Recht reguliert Beziehungen, schreibt im Detail aber nur in Ausnahmefällen vor, was moralisch geboten ist. Eine moralische Haltung lässt sich nicht aus der Rechtsförmigkeit unseres Handelns ableiten. Rechtliche Rahmungen lassen Raum für unmoralisches Verhalten. Ethik hat einen transzendierenden Charakter, der bei der Bewertung von Technologien, die ein Veränderungspotential in Bezug auf unser Selbst- und Gesellschaftsverständnis haben, eine besondere Rolle spielt. Moralisches Handeln deckt sich nicht notwendig mit den Vorgaben des positiven Rechts. Zwar sind Kompetenzzuweisungen zwischen Chirurg und OP-Schwester geregelt, es sind aber Situationen denkbar, in denen eine

Kompetenzüberschreitung moralisch geboten, nicht aber rechtlich gedeckt werden kann. Gesetze müssen Spiel haben, um dauerhaft anwendbar zu sein - dies gilt besonders bei Regelungen, die technische Entwicklungen betreffen. Der ethische Diskurs hat eine motivierende Komponente, da er begründet, warum ein bestimmtes Handeln geboten ist. Der ethische Diskurs geht also 1) nicht im rechtlichen auf, weil er auch eine kritische, die bestehende Moral überprüfende sowie transzendierende Aufgabe hat; er beschäftigt sich 2) nicht mit Regulierungen, die ohne unmittelbaren moralischen Bezug sind; und er hat 3) einen motivierenden Charakter, der das Überdenken bestehender rechtlicher Regelungen einschließt.

Leitend für die Bewertung von Big-Data-Anwendungen sind 1) metaethische Fragen, die die formale Grundlage des ethischen Diskurses betreffen, 2) ethisch-normative Fragen, die den Diskurs in einem materialen Sinne charakterisieren und 3) die Frage nach der Entscheidung, denn bei der Bewertung solcher Anwendungen spielt die Zuschreibung einer Entscheidungskompetenz eine wichtige Rolle.

Metaethische Fragen sind 1) die Frage nach der Identität des Handlungssubjekts, dem verantwortliches Handeln zugeschrieben werden soll; 2) die nach der Bestimmung der Wirklichkeit, in der gehandelt werden soll; 3) die nach der Wahl, denn wir können nur verantworten, was wir auch gewählt haben. Konkret heißt das: Wird die Identität des Handlungssubjektes durch die Nutzung einer Technik geschwächt, infrage gestellt oder aufgehoben, wird die Wirklichkeitsbestimmung für das Subjekt, das handeln soll, erschwert oder verunmöglicht, und wird Wahl reduziert oder gar aufgehoben.

Normative Fragen leiten den ethischen Diskurs inhaltlich. Dies trifft auf Leitwerte des Selbst- und Gesellschaftsverständnisses zu, wie sie im Grundgesetz als Vermächtniswerte aufgegeben sind. Dies ist 1) die Würde als unveräußerliches Recht des Menschen als selbstzweckhaftes,

einzigartiges Wesen behandelt, also nie nur in seiner Rolle wahrgenommen zu werden; 2) die Autonomie des Einzelnen, der das Recht hat, sein eigenes Leben zu führen; 3) die Subsidiarität als Sicherung gegen Entmündigung und Paternalismus sowie als Prinzip der Gesellschaftsorganisation, nach dem als Verantwortungsträger möglichst alle zu beteiligen sind und Macht nur mit Verteilung auf viele legitimiert ist. Das Subsidiaritätsprinzip soll verhindern, dass ein autonomes Wesen bevormundet wird, selbst wenn dies den Interessen der Gesellschaft entgegenkommt. Ihm darf aber auch nichts abgenommen werden, was es selbst entscheiden und leisten kann. Das Subsidiaritätsprinzip verhindert zwar Entmündigung, bürdet aber auch Verantwortung auf.

Zuletzt ist die Frage nach der ‚Entscheidung‘ zu fokussieren, die die für den ethischen Diskurs zentrale Grundannahme des Handelns betrifft. Von Handeln können wir reden, wenn Folgenverantwortlichkeit, Mittelwahlkompetenz und Zwecksetzungsautonomie vorliegen. Billigen wir Systemen Entscheidungsfähigkeit zu, müssen wir ihnen auch Handlungsfähigkeit zuschreiben, was problematisch ist, solange sie unsere Werkzeuge sind. Während ihnen Mittelwahlkompetenz zugesprochen werden könnte, wäre Folgenverantwortlichkeit nur denkbar, wenn sie für ihre Entscheidung zur Rechenschaft gezogen werden könnten, und Zwecksetzungsautonomie nur, wenn sie eigene Zwecke verfolgten. Ein System würde dann entscheiden, wenn es über eine hermeneutische Kompetenz verfügt, in der es sich gegenüber dem Ausgelegten positioniert. Die Rede von Systementscheidung verschleiert die Tatsache, dass es nicht Systeme sind, die für bestimmte Rechenergebnisse verantwortlich sind, sondern diejenigen, die Algorithmen entwickeln, implementieren und die Ergebnisse der Systeme anerkennen.

Im Big-Data-Diskurs fällt auf, dass informationswissenschaftliche Unterscheidungen, wie die von Datum, Information und Wissen, wenig

präsent sind und die Diskussion Unschärfen aufweist, die für Fehleinschätzungen verantwortlich sind. In der Informationswissenschaft dominiert ein Begriffsgebrauch, der Daten als ‚semantikfreie‘ Gegebenheiten versteht und Informationen als erkannte und bewertete. Wissen dagegen resultiert aus der Relationierung und Hierarchisierung von Informationen. Daten, Informationen und Wissen stehen in einem selbstreferentiellen Verhältnis und sind nicht isoliert diskutierbar. Um die Debatte über Daten-, Informations-, und Wissensgesellschaft zu verstehen, lohnt ein Blick auf die Begriffsgeschichte.

Der lateinische Begriff ‚datum‘ wurde von Chr. Wolff im 17. Jh. mit Gegebenes übersetzt, womit ein individuell Seiendes, das sich dem Rezipienten ohne dessen Zutun zeigt, als Voraussetzung der Wahrnehmung genommen wird. In der Folge wurde einerseits bestritten, dass es sich bei Daten um einen Rohstoff der Wahrnehmung handelt und deren Konstruiertheit angenommen; andererseits setzte sich in positivistischen Strömungen ein Gebrauch durch, der Daten an unsere Empfindungen knüpft und als Tatsachen versteht.

Der aus dem Lateinischen ‚informare‘ abgeleitete Begriff vereint eine direkte Bedeutung: ‚formen‘ sowie eine übertragene: ‚unterrichten‘. Infolge der Verwendung des Begriffes in Nachrichtentechnik und Kybernetik rückt die semantische und pragmatische Bedeutung in den Hinter- und eine formal-syntaktische in den Vordergrund. Shannon hat den Begriff aus dem menschlichen Gebrauchskontext genommen und als quantifizierbare Größe aufgefasst, die von Semantik und interpretativen Aspekten absieht. Mit Information wird die Wahrscheinlichkeit des Eintretens eines Ereignisses bezeichnet. Sie verringert die Ungewissheit und erhöht die Ordnung eines Systems. Im sprach- und informationswissenschaftlichen Begriff werden dagegen semantische und pragmatische Aspekte sowie die subjektive Wissensveränderung betont.

Die umfangreichste Geschichte hat der Begriff des Wissens. Dabei wurde das wissenschaftliche Wissen, das sich durch Begründung, Kohärenz, Wahrheitsbezug und intersubjektive Anerkennung auszeichnet, als zentraler Wissenstyp herausgearbeitet. Es geht bei ihm um gerechtfertigtes, intersubjektiv ausweisbares Wissen, was die Griechen ἐπιστήμη nannten und mit Wissenschaft gleichsetzten. Es ist durch Reflexion gewonnen und entsteht aus unserer Fähigkeit zur Zuordnung, Hierarchisierung und Selektion. In Bezug auf die Diskussion der Nutzung informatischer Werkzeuge impliziert es die Reflexion der Systeme, die Ordnungs- und Einbindungsfunktionen für Informationen übernehmen.

Reflexionswissen leistet nicht nur Welt-, sondern auch Selbsterkenntnis. Als Orientierungswissen ist es auch als hermeneutisch zu bezeichnen, denn es befähigt uns Weltzusammenhänge und uns selbst zu verstehen. Es setzt kontextuell gebundenes Vor-Wissen voraus, kommt nicht voraussetzungslos zustande und ist erweiter- und korrigierbar.

An die Stelle der Rede von der Informationsgesellschaft als Signatur unserer Zeit ist seit Ende der 1990er Jahre der Begriff der Wissensgesellschaft getreten. H. Markl bestimmt letztere als eine Konsequenz aus der Informationsgesellschaft, die sich auf einen hermeneutischen und wertenden Weg begeben muss, will sie nicht in Bedeutungslosigkeit versinken. Damit schien sich eine Abkehr von semantikfreien informationstheoretischen Modellen anzukündigen.

In dem gegenwärtigen Big Data-Diskurs ist nicht nur ein Rückfall von der Wissens- in die Informationsgesellschaft zu konstatieren, sondern eine Unterbietung der überwunden geglaubten Informations- durch die Datengesellschaft. Implizit wird damit der Anspruch erhoben, dass die Rahmung von Daten und semantische Zuordnungen keine Relevanz mehr hätten. Es wird davon ausgegangen, dass es sich bei Daten um einen Rohstoff handelt, aus dem Welt beliebig ‚zusammengesetzt‘ werden

könne. Es werden Korrelationen erkannt und Analyse unmittelbar für Steuerungsprozesse genutzt. Mit Big-Data-Technologien ist die überwunden geglaubte ‚Massenideologie‘ wieder in den Fokus gerückt.

Daten sind kein Rohstoff, keine Vorgegebenheiten im Sinne der antiken ‚physis‘, die sich uns unabweisbar aufdrängt, sondern Ergebnisse von Erfassungs- und Sammelprozessen. Die Vorstellung von der Semantikfreiheit der Daten ist irreführend da sie Ergebnis vorgängiger Bewertungsprozesse sind. Wenn Daten im Unterschied zu bereits bewerteten Informationen oder zu handlungsrelevantem Wissen in unterschiedliche Kontexte gesetzt werden können, bedeutet dies nicht, dass sie denselben Status wie die antike ‚physis‘ hätten. Informatische Daten kommen nicht von selbst zustande, drängen sich uns nicht auf und wir können uns ihnen entziehen. Sie sind Ergebnis eines technisch-praktischen Zugriffs auf die Welt, markieren ein Verhältnis von Subjekt und intendiertem Objekt und kommen durch Relevanzzuschreibung zustande. Bei jeder Messung ist der Wert der Messung durch die Intention festgelegt, die mit ihr verfolgt wird. Informatische Probleme bei der Bewältigung großer Datenmengen rühren nicht zuletzt daher, dass diese etwas Präformiertes sind. Ein Datum verweist auf Mitdaten und eine gemeinsame Rahmung. Hinter der Datafizierungs-idee steht eine Metaphysik, die die Vermitteltheit von Daten in Bewertungs- und Selektionsprozessen unterschlägt.

Wie wirkt sich nun der Einsatz von Big-Data-Technologien auf unser Selbst, Gesellschafts- und Weltverständnis aus? Letzteres wandelt sich insofern, als wir zunehmend davon ausgehen, dass die Welt in einem beliebig manipulierbaren digitalen Double zu erfassen sei. Wirklichkeit in ihrer Widerständigkeit und Fülle ist aber nicht zu verdoppeln. Selbst künstliche Organe sind keine Verdoppelung des ausgefallenen oder geschädigten Organs, sondern ein Modell der als zentral erachteten Funktionen.

Wo menschliche Verantwortung hinter dem System verschwindet, muss von einer Ideologisierung der Technologie gesprochen werden. Problematisch ist, wenn das Individuelle und Ereignishafte aus dem Auge verloren wird. Es kann zu einer Entethisierung kommen, wenn suggeriert wird, dass historische Ereignisse oder personales Verhalten berechenbar, Wirklichkeit die Summe ihrer Daten und Fehlprognosen allein auf einen Datenmangel zurückzuführen seien. Wirklichkeit ist aber nicht die Summe der Daten, die wir über sie haben, weil wir sie perspektivisch und diskursiv fassen und sie unterschiedlich artikulieren. Wirklichkeitserfahrung artikuliert sich vielmehr in Widerständigkeitserfahrungen gegenüber unserem Formwillen. Wo nur gerechnet wird, wird nicht entschieden und damit auch nicht gehandelt.

Nicht das System verantwortet seine Entscheidung, sondern diejenigen die die Entscheidung des Systems disponieren und anerkennen. Anerkennungsverhältnisse können aber auch manipulativ hergestellt werden, indem man Vorteile durch die Systemnutzung verspricht oder bei Nichtanerkennung mit Nachteilen oder Diskursausschluss droht. Zuletzt können Anerkennungsverhältnisse durch eine ‚Überzeugungsmacht‘ hergestellt werden, die sich nicht kraft des besseren Arguments durchsetzt, sondern kraft besserer Sichtbarkeit. Überzeugungsmacht hat etwas mit Selektionsmacht und der Fähigkeit zu tun, widerstreitende Argumente unsichtbar zu machen. Wird die historische Vermittlung und perspektivische Erfassung der Wirklichkeit ausgeblendet, haben wir es mit einer Ideologisierung des scheinbar Bestehenden zu tun.