Technische Universität Berlin

Wintersemester 2020

Fakultät I: Geistes- und Bildungswissenschaften

Institut für Sprache und Kommunikation

Fachgebiet: Medienwissenschaft

Masterarbeit für die Prüfung zum Master of Arts im Studiengang Medienwissenschaft

Erstgutachter: Prof. Dr. Stephan Günzel

Zweitgutachter: Dr. Adina Lauenburger

Unboxing the Black Box – Gesellschaftliche Implikationen algorithmischer Entscheidungsfindung

Abgabetermin: 20.09.2020

Eingereicht am: 18.09.2020

Vorgelegt von: Hans Stiegler

Matrikelnummer: 384193

Masterstudiengang Medienwissenschaft

E-Mail: h.m.stiegler@campus.tu-berlin.de

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den

Unterschrift

Inhaltsverzeichnis

1 Einleitung

1.1 Problemstellung und Motivation

Ideen und Vorstellungen von Künstlicher Intelligenz haben Menschen bereits seit der Antike fasziniert. Sei es in Form der altgriechischen Pandora, die von den Göttern geschaffen wurde um auf Erden ihre Büchse zu öffnen und so die Menschheit zu strafen (vgl. Mayor 2018), in Form des biblischen Golems (vgl. Tinnefeld 2019) oder in den vielen Variationen anthropomorpher Intelligenzen moderner Science-Fiction-Geschichten.

Mittlerweile ist Künstliche Intelligenz mitten im gesellschaftlichen Leben angekommen. Allerdings nicht als rebellierender, gesprächiger Bordcomputer wie HAL aus *2001: Odyssee im Weltraum* oder als menschenähnlicher Roboter, sondern als breites Spektrum abstrakter Algorithmen, die oftmals ohne viel öffentliche Diskussion fast jeden Aspekt des gesellschaftlichen Lebens beeinflussen. Algorithmen berechnen die Höhe von Versicherungsprämien und entscheiden über die Kreditwürdigkeit von Antragstellern anhand ihrer demographischen und persönlichen Daten oder auf Basis ihres Verhaltens (Welchering 2018). Im medizinischen Bereich werden KI-Systeme eingesetzt und schlagen durch überlegene Mustererkennung bereits menschliche Ärzte bei der Diagnose von Hautkrebs (Brinker et al. 2019). Das US-Verteidigungsministerium hat eine “Algorithmic Warfare”-Gruppe eingerichtet und plant, eine Milliarde US-Dollar für KI auszugeben. Neben dem Einsatz von KI in bewaffneten Robotern und in der Datenverarbeitung sollen algorithmische Systeme entwickelt werden, die feindliche Truppenbewegungen vorhersagen können und so vormals exklusiv menschliche Entscheidungen mitbestimmen (The Economist 2019). Auch in Deutschland setzt die Polizei im Rahmen des “Predictive Policing” Algorithmen zur Unterstützung der Polizeiarbeit ein, zum Beispiel um mit dem System “SKALA” Wohnungseinbrüche in nordrhein-westfälischen Großstädten vorherzusagen (LKA Nordrhein-Westfalen 2018).

Diese Systeme, die große Mengen Daten nutzen um Vorhersagen zu treffen und damit Entscheidungsfindungen zu unterstützen lassen sich unter dem Begriff des “Algorithmic

Decision-Making” (ADM) zusammenfassen. Der Einsatz dieser Systeme durch Unternehmen, Behörden, Universitäten und vielen weiteren gesellschaftlichen Institutionen stellt nicht weniger als einen Paradigmenwechsel dar, der grundsätzliche technische, ethische und gesellschaftliche Fragen aufwirft. So sieht es auch Bundeskanzlerin Angela Merkel, wenn sie betont: “Es sind sich alle Beteiligten einig, dass es sich hier um gewaltige Veränderungen unserer Lebens- und Arbeitswelt handelt und dass diese Veränderungen den Menschen zu dienen haben und nicht etwa die Technik nur als solches betrachtet werden kann” (Absatzwirtschaft 2019).

Kritik und Skepsis an dem Einsatz von Künstlicher Intelligenz und algorithmischer Entscheidungsfindung ist nicht neu, sondern wurde bereits formuliert bevor neuere technologische Entwicklungen den flächendeckenden Einsatz dieser Systeme überhaupt möglich gemacht haben. Joseph Weizenbaum, einer der Pioniere der KI-Forschung im 20. Jahrhundert, diskutiert in seinem 1976 veröffentlichten Buch “Computer Power and Human Reason” (deutscher Titel: “Die Macht der Computer und die Ohnmacht der Vernunft”) grundsätzliche ethische und philosophische Fragen, die durch Entwicklungen in der KI-Forschung aufgeworfen werden. In den 1960er-Jahren entwickelte Weizenbaum am MIT das Programm “ELIZA”; ein Computerprogramm, das ein Gespräch mit einem Therapeuten nach dem Vorbild der Rogerschen Gesprächstherapie simulierte. Obwohl simpel nach heutigen Maßstäben, beeindruckte ELIZA 1966 Wissenschaftler und Nutzer. Testpersonen wollten alleine sein um eine intime Gesprächssituation mit dem Computer herzustellen und Psychiater stellten sich vor, ELIZA flächendeckend für therapeutische Maßnahmen einzusetzen. Weizenbaum selbst hingegen war schockiert von den positiven Reaktionen auf seine Kreation und hielt sie für naiv, gefährlich und geleitet von einem falschen Menschenbild, das den Menschen als nicht mehr als eine organische Rechenmaschine betrachtete.

Während sich viele seiner zeitgenössischen KI-Kollegen aus Wissenschaft und Praxis danach mit Fragen der technischen Machbarkeit von KI beschäftigten, sah Weizenbaum die zentrale Frage im Konflikt zwischen Mensch und Maschine. Anstatt darüber zu diskutieren ob Computer jemals schnell genug und Algorithmen jemals effizient genug sein werden um wichtige Entscheidungen zu treffen, stellte Weizenbaum eine

grundsätzlichere Frage: welche Entscheidungen sollten überhaupt von Computern getroffen werden dürfen? Für Weizenbaum war klar, dass bestimmte Entscheidungen und Rollen so wichtig sind, dass sie niemals von Maschinen übernommen werden sollten. Zentral ist dabei sein Argument, dass Menschen eben keine organischen Rechenmaschinen sind und deswegen nicht von Maschinen ersetzt oder beurteilt werden sollten. Weizenbaum sieht diese weit verbreitete Position im Computer als Medium und Werkzeug verkörpert, beziehungsweise in der “Computer-Metapher”, dessen inhärente Logik das Denken über KI und ihren Einsatzmöglichkeiten maßgeblich beeinflusst und leitet. Der im Kern medientheoretische Ansatz von Weizenbaum ermöglicht eine Betrachtung von KI-Problemen nicht bloß als lose Serie technischer Probleme die mit schnelleren Rechnern und besseren Methoden gelöst werden können, sondern macht eine fundamentale Analyse von künstlicher Intelligenz und ihrem gesellschaftlichen Einfluss möglich.

1.2 Zielstellung und Aufbau der Arbeit

Das Ziel dieser Arbeit ist es, Weizenbaums Überlegungen aus “Computer Power and Human Reason” in einen medientheoretischen Kontext zu setzen um so die Frage nach den gesellschaftlichen Implikationen moderner KI-Technologien und ihren Anwendungsformen stellen zu können. Aus der Verbindung von Weizenbaums medientheoretischen Überlegungen mit praktischen Problemen aus dem Bereich der algorithmischen Entscheidungsfindung aus den letzten Jahren ergibt sich die folgende Forschungsfrage: *ist Joseph Weizenbaums Kritik an KI-Technologien aus dem Jahr 1976 auch im Jahr 2020 noch relevant?*

In Kapitel 2 soll zuerst auf Joseph Weizenbaums biographischen Kontext und die zentralen Überlegungen seines Buchs “Computer Power and Human Reason” (1976) eingegangen werden. Weizenbaum beschäftigt sich in “Computer Power and Human Reason” mit einer Vielzahl unterschiedlicher Themen. Ziel dieses Kapitels ist es, medientheoretische Positionen herauszuarbeiten und in einen medienwissenschaftlichen Kontext einzubinden.

Kapitel 3 beschäftigt sich mit mit Definitionen von künstlicher Intelligenz, ihrer historischen Entwicklung und ihrer technischen Implementierungen. In den 2010er-Jahren gab es innerhalb der KI-Forschung und -Praxis eine Trendwende von regelbasierten, linearen Methoden hin zu datenbasierten, non-linearen Methoden wie Machine Learning und Neural Networks. Dieser technische und methodische Paradigmenwechsel hat einen großen Einfluss auf die Entwicklung und Anwendung von KI-Technologien, der in diesem Kapitel dargestellt wird.

Kapitel 4 beschäftigt sich mit algorithmischer Entscheidungsfindung (Algorithmic Decision-Making, kurz: ADM), einem Anwendungsgebiet von Künstlicher Intelligenz mit massiven gesellschaftlichen Implikationen. Neben einer Übersicht, wo und wie ADM-Systeme eingesetzt werden, wird auf den gesellschaftlichen Einfluss dieser Systeme eingegangen. Algorithmische Entscheidungsfindung im gesellschaftlichen Kontext wirft spezifische Probleme auf, auf die gemeinsam mit möglichen Lösungsansätzen eingegangen werden soll.

In Kapitel 5 werden die Ergebnisse dieser Arbeit schlussendlich zusammengefasst und, in Rückblick auf die Forschungsfrage, ein Fazit gezogen das Weizenbaums Kritik aus den Frühzeiten der KI-Forschung in einen aktuellen Kontext einordnet. Zusätzlich wird auf weiteren Forschungsbedarf eingegangen und offen gebliebene Fragen diskutiert.

2 Joseph Weizenbaums KI-Philosophie

2.1 Biographischer Überblick

Joseph Weizenbaum wurde 1923 in Berlin geboren und verließ 1936 nach der Machtübernahme der Nationalsozialisten Deutschland und emigrierte mit seiner Familie in die Vereinigten Staaten. Nach einem abgebrochenen Studium der Mathematik in Detroit arbeitete er als Meteorologe für die US-Armee und schloss nach dem Ende des zweiten Weltkriegs sein Mathematik-Studium ab. Später arbeitete er als Informatiker für Bank of America, bis er 1962 erst als Gastprofessor, dann ab 1970 als Professor für Informatik, zum Massachusetts Institute of Technology (MIT) wechselte. 1996 kehrte er nach Deutschland zurück, wo er 2008 im Alter von 85 in Gröben bei Berlin verstarb (Markoff 2008). Weizenbaum befasste sich während seiner Laufbahn intensiv mit den sozialen und ethischen Implikationen seiner Arbeit als Informatik-Professor und betrachtete den Computer äußerst kritisch, weswegen er sich innerhalb der KI-Forschung und der breiteren Wissenschaftsgemeinde selbst als “Häretiker” bezeichnete (Long 1985). Diese Position hatte er nicht zuletzt wegen seiner Herkunft eingenommen. Bis zum dreizehnten Lebensjahr in Deutschland aufgewachsen, musste er als Jugendlicher nach der Machtergreifung der Nationalsozialisten mit seiner Familie in die USA flüchten. Sein Wissen darüber, dass große Teile der deutschen Wissenschaftsgemeinde opportunistisch unter den Nationalsozialisten weitergearbeitet haben, brachte ihn zu der Auffassung dass Wissenschaftler sich nicht nur rein fachlich betätigen sollte, sondern auch eine Pflicht haben, politische und soziale Fragen zu stellen.

Die verbreitete Annahme seiner Zeit, dass der Computer grundsätzlich in jedem Bereich eine Verwendung finden könnte, hielt Weizenbaum für falsch. Er sah den Computer als “Lösung, die nach einem Problem sucht”: Seiner Auffassung nach wird die Frage nach dem Einsatz von Computern rückwärts gestellt. Anstatt davon auszugehen, dass der Computer in jedem Fall eine Unterstützung sein kann und nach Problemen zu suchen die mit dem Computer gelöst werden können, sollte grundsätzlich geklärt werden, was die Ziele sind und wie diese erreicht werden können – und ob der Computer dafür das

geeignete Werkzeug ist. Erst nach dieser Analyse ist es sinnvoll, zu untersuchen ob diese Ziele mit Hilfe des Computers gelöst werden können.

Den Computer selbst begann er schon früh als “fundamental konservative Kraft” zu sehen, die technische Lösungen möglich macht, wo sonst nach sozialen Lösungen hätte gesucht werden müssen. Fast paradoxerweise blockiert der Computer durch seine enorme Effektivität somit soziale Fortschritte und Veränderungen des herrschenden Systems, wodurch bestehende Machtstrukturen bestehen bleiben oder sogar verstärkt werden. Die technische Lösung durch den Computer ist aber wenig mehr als ein Pflaster und keine tatsächliche Lösung des Problems. Dies hängt für Weizenbaum auch damit zusammen, dass der Computer zuerst im militärischen Bereich entwickelt wurde und deswegen von Anfang an ein “Militärinstrument” gewesen sei (vgl. ben-Aaron 1985).

2.2 ELIZA

Weizenbaums Reflektion über die sozialen Implikationen künstlicher Intelligenz begannen mit den Reaktionen auf sein Computerprogramm “ELIZA”, benannt nach der Figur Eliza Doolittle aus Bernhard Shaws’ Theaterstück “Pygmalion”. ELIZA wurde von Weizenbaum ab 1964 in den ersten Jahren seiner Gastprofessur am MIT entwickelt und löste die erste Welle von Interesse an künstlichen Konversations-Systemen aus (Wallace 2008).

Testpersonen konnten eine einfache Konversation mit dem Programm führen, allerdings hatte ELIZA zu keinem Zeitpunkt ein semantisches Verständnis von den analysierten Texten oder produzierten Antworten und auch nicht das Ziel, eine solche Art von Intelligenz zu erreichen. Die Interaktion zwischen dem Computerprogramm und dem eingegeben Text war rein symbolisch; Weizenbaum wollte mit ELIZA vielmehr die Lösung mehrerer technischer Probleme darstellen. Dazu gehörten die Erfassung wichtiger Schlüsselwörter und des Kontexts in dem diese Wörter auftauchen, ihrer Beziehungen untereinander als Verben, Objekte und Subjekte, sowie der Umgang mit Texten die keine erkennbaren Schlüsselwörter enthalten. Außerdem sollte dargestellt werden, wie auf Basis dieser Textanalyse ein Skript dynamisch angepasst und wiedergegeben werden kann

(Weizenbaum 1966). Damit stellt ELIZA keine Intelligenz im menschlichen Sinne dar, sondern vielmehr eine Maschine zur Mustererkennung (Block 1981).

Benutzer von ELIZA konnten am Computer einen Text eintippen, der von ELIZA analysiert und mit offenen, weiterführenden Fragen beantwortet wurde. Im Hintergrund bestand das Programm aus zwei Komponenten: einer Analyse- und einer Skript-Komponente. Die Analyse-Komponente empfängt den eingehenden Text, analysiert ihn nach allgemeinen Regeln und bricht ihn in Bestandteile wie Schlüsselwörter herunter, die für die Skript-Komponente verständlich sind. Die austauschbare Skript-Komponente übernimmt ab diesem Punkt und besteht aus einer Reihe von Regeln und möglichen Antworten auf den eingehenden Input. Das bekannteste ELIZA-Skript wurde “DOCTOR”; dieses Skript ermöglichte es ELIZA, offene Fragen im Stil der psychotherapeutischen Gesprächsmethodik des Psychologen Carl Rogers zu stellen. Dieses Skript wurde von Weizenbaum gewählt, da der antwortende, psychiatrische Konversationspartner nichts über die reale Welt wissen muss aber dessen auffordernde Antworten vom sprechenden Partner als intelligent interpretiert wird. Die offene Frage “Erzählen Sie mir mehr über…” wird vom sprechenden Partner nicht als Unwissenheit interpretiert, sondern als vermeintlich zielgerichtet auf eine tiefere Bedeutung (vgl. Weizenbaum 1966, 6). Diese Art von Konversation ließ sich dementsprechend relativ leicht umsetzen und erlaubte ELIZAs Mustererkennung “intelligent” zu wirken, ohne weiterführendes Wissen über die getätigten Aussagen des Sprechers zu haben (vgl. Weizenbaum 1976, 3-4).

Obwohl Weizenbaum selbst seine in ELIZA implementierte Version der Rogerschen Gesprächsführung eher als “Parodie” denn als ernstgemeinte Simulation einer Konversation sah (vgl. Weizenbaum 1976, 188), gelang ELIZA nach seiner Veröffentlichung 1966 als vermeintlich intelligentes Computerprogramm schnell zu Berühmtheit und wurde einem breiten Publikum bekannt. Wie die Öffentlichkeit auf ELIZA reagierte überraschte Weizenbaum nicht nur, sondern schockierte ihn. Er beschreibt drei spezifische Reaktionen auf ELIZA, die ihn zum weiteren Nachdenken über die Rolle des Computers in der Gesellschaft nachdenken ließen. Bereits während Weizenbaum ELIZA mit echten Menschen ausprobierte fiel ihm auf, dass die Testperson innerhalb kürzester Zeit begannen, mit ELIZA wie mit einem anderen Menschen zu kommunizieren und eine

intime, quasi-soziale Verbindung aufzubauen. So bat Weizenbaums Sekretärin ihn, während ihrer “Konversation” mit ELIZA den Raum zu verlassen, um ihr privates Gespräch nicht zu stören. Weitere Testpersonen protestierten gegen die Aufzeichnung und Auswertung ihrer Gespräche mit ELIZA, als ob diese Aufzeichnungen das Produkt einer realen Konversation wären. Insbesondere die Schnelligkeit, mit der seine Testpersonen eine parasoziale Beziehung mit dem, im Kern äußerst simplen, Programm aufbauten, überraschte Weizenbaum (vgl. Weizenbaum 1976, 6-7). Neben den Reaktionen seiner Testpersonen fand Weizenbaum auch die Reaktion aus dem Gesundheitswesen bemerkenswert. Psychiater reagierten äußerst positiv auf ELIZAs “DOCTOR”-Skript und sahen schnell Einsatzmöglichkeiten für diese Technologie im therapeutischen Bereich. Mit einem Programm wie ELIZA könnten beispielsweise psychotherapeutische Maßnahmen und Gespräche zwischen Therapeut und Patient zu einem großen Grad automatisiert werden. Weizenbaum sah in diesen Ideen eine schockierende Tendenz, den therapeutischen Prozess auf eine simple Interviewmethode zu reduzieren und andere Elemente, wie das Bedürfnis des Patienten nach einem empathischen menschlichen Zuhörer, außer Acht zu lassen (vgl. Weizenbaum 1976, 5-6). Große Teile der Öffentlichkeit die von ELIZA hörten, verstanden ELIZA als allgemeine Lösung für die mechanische Verarbeitung von Sprache, obwohl Weizenbaum im ELIZA-Paper explizit darauf hinweist, dass ELIZA weder diese Lösung anbietet, noch den Anspruch hat eine solche Lösung anzubieten. Da Sprache immer in ein spezifischen Kontext eingebettet ist, kann es eine solche allgemeine Lösung nicht geben – selbst Menschen stellen nach Weizenbaum keine solche allgemeine Lösung dar, da auch unterschiedliche Personen in einer Unterhaltung nie über den exakt gleichen sprachlichen Kontext verfügen (vgl. Weizenbaum 1976, 7-8).

Diese Reaktionen auf ELIZA zeigten Weizenbaum, dass vermeintlich intelligente Technologien wie ELIZA einen enormen Einfluss auf die Wahrnehmung und das Denken von Menschen haben können. Schockiert von den Reaktionen befasste Weizenbaum sich mit den Fragen, warum Menschen so schnell mit einem simplen Programm eine Beziehung eingehen; warum selbst erfahrene praktizierende Psychiater den therapeutischen Prozess auf eine regelbasierte Konversation reduzieren und welche Rolle Wissenschaftler in der Vermittlung von Wissen spielen, wenn selbst ein simples Programm wie ELIZA von der breiten Öffentlichkeit als intelligente Lösung von

Sprachverarbeitung aufgenommen wird. In allen Reaktionen sah Weizenbaum einen gemeinsamen Ursprung: die weitverbreitete “mechanische Idee des Menschen” (“mechanical conception of man”) (vgl. Weizenbaum 1976, 1). Dieser reduktionistischen Idee nach ist der Mensch nichts anderes als eine dem Computer ähnliche organische Datenverarbeitungsmaschine. Nur unter dieser Annahme, so Weizenbaum, können Psychiater ihre Arbeit als rein regelbasiert betrachten, seine Testpersonen die Maschine wie einen Computer behandeln oder annehmen, dass ein Computer Sprache verstehen kann.

Diese Reaktionen inspirierten Weizenbaum letztendlich dazu “Computer Power and Human Reason” (1976) (Deutsch: “Die Macht der Computer und die Ohnmacht der Vernunft”) zu schreiben um sich ausführlicher mit den Fragen auseinanderzusetzen, die sich ihm nach der Veröffentlichung von ELIZA stellten.

2.3 Computer Power and Human Reason

In “Computer Power and Human Reason” (1976) stellt Weizenbaum den Konflikt zwischen Mensch und Maschine zentral. Wie der Titel des Buchs bereits verrät, sieht Weizenbaum die Rechenleistung des Computers als Gegenspieler der menschlichen Vernunft. Der deutsche Titel “Die Macht der Computer und die Ohnmacht der Vernunft” macht seine zugrundeliegende Perspektive noch deutlicher: die menschliche Vernunft verliert gegenüber der quantitativen Kraft des Computers. In dem Buch beschäftigt Weizenbaum sich mit zwei Fragen: dem Unterschied zwischen Mensch und Maschine und, ausgehend von diesem zentralen Konflikt, mit der Frage nach den Grenzen des Computers. Nach Weizenbaum stellt sich dabei nicht die Frage nach den technischen Grenzen und Möglichkeiten des Computers, sondern ob der Computer überhaupt für bestimmte Zwecke eingesetzt werden sollte.

Bei dieser Auseinandersetzung stellt er den Gedanken zentral, dass die Welt vom Menschen bereits zu sehr zu einem Computer, also quantitativ-reduktionistisch, gemacht wurde und bestimmte Perspektiven auf die Welt nicht erst durch den Computer entstanden sind, sondern schon vorher existierten. Der Computer als moderne

Universalmaschine stellt für ihn vor allem eine Verkörperung dieser Ideen dar, die durch ihn erst besonders deutlich hervorgetreten sind: “In an important sense, the computer is used here merely as a vehicle for moving certain ideas that are much more important than computers [...] But a major point of this book is that we, all of us, have made the world too much into a computer [...]” (vgl. Weizenbaum 1976, x-xii). Im folgenden sollen die Kernthemen und -positionen von Weizenbaums “Computer Power and Human Reason” dargestellt werden: die Signifikanz des Computers als Medium beziehungsweise Werkzeug, die Ideen die im Computer ihren Ausdruck finden, die Unterschiede zwischen Mensch und Maschine sowie die aus diesen Unterschieden resultierenden Grenzen des gesellschaftlichen Einsatzes von Computern. Dies stellt die Basis für eine spätere medientheoretische Einordnung von Weizenbaums Ideen dar und beschreibt, wie Weizenbaum die gesellschaftliche Rolle des Computers sieht und welche ethischen Implikationen diese Rolle aufwirft.

2.3.1 Der Computer als Werkzeug

Werkzeuge, Technologien und Methoden spielen kulturgeschichtlich eine entscheidende Rolle in der Entwicklung des Menschen. Die Werkzeuge die der Mensch benutzt und die Beziehung zu diesen Werkzeugen bestimmen, wie der Mensch seine Umwelt versteht und betrachtet und damit auch sich selbst. Da Menschen sich etwas erst vorstellen müssen, bevor sie es umsetzen können, stellt das Verhältnis zwischen dem Mensch und seinen Werkzeugen eine eine komplexe Wechselbeziehung dar zwischen dem, was Menschen möglich ist und was sie für möglich halten. Der Mensch nutzt Werkzeuge, um seine materielle Umwelt nach seinen Vorstellungen zu verändern; gleichzeitig erweitern Werkzeuge und ihre Möglichkeiten die Vorstellungskraft des Menschen und schaffen damit neue Einsatzmöglichkeiten. Innerhalb der menschlichen Vorstellungskraft entsteht so ein subjektives Modell der Welt und des Menschen das signifikant geprägt ist durch die zur Verfügung stehenden Werkzeuge. Wenn Menschen Entscheidungen treffen und mit ihrer Umwelt interagieren, ist es dieses subjektive Modell der äußeren Welt mit dem sie interagieren und nicht mit einer objektiven Realität. Innerhalb dieser subjektiven Realität sind Mensch und Werkzeug so verwoben, dass eine Trennung des Menschenbilds und den genutzten Werkzeugen unmöglich ist (vgl. Weizenbaum 1976, 17-18).

Weizenbaum unterscheidet zwischen “prosthetischen” Werkzeugen, die die natürlichen Fähigkeiten des Menschen erweitern und “autonomen” Maschinen, die selbständig nach den Regeln einer inneren Logik arbeiten, welche einen bestimmten Teil der Welt modelliert. Auch prosthetische Werkzeuge entwickeln die Vorstellungskraft des Menschen weiter; Waffen ermöglichten eine effizientere Jagd und erweiterten damit die Möglichkeiten des Menschen sich auszubreiten, genauso wie warme Kleidung, mit der sich Menschen neue Territorien und Klimazonen erschließen konnten. Während prosthetische Werkzeuge somit eine Art natürliche Erweiterung und Verlängerung menschlichen Handelns darstellen, erzeugen autonome Maschinen eine kategorisch andere Interaktion zwischen Mensch und Umwelt. Bis zur Erfindung des Computers war, so Weizenbaum, die Uhr die einzige wirklich wichtige autonome Maschine. Die Uhr ist autonom, da innerhalb der Uhr ist ein bestimmtes Modell von Zeit kodiert ist, das unabhängig von menschlicher Wahrnehmung funktioniert. Während Menschen vormals Zeit am Stand der Sonne oder dem Krähen eines Hahns festgemacht haben, begann mit der Uhr die Idee von Zeit als messbare Größe von diskreten, aufeinanderfolgenden Einheiten. Die Benennung dieser Einheiten als Sekunden, Minuten, Stunden, und so weiter, macht diese vormals nicht existierende Einteilung noch realer und verstärkt sie somit. Durch die Anwendung dieser Einteilung auf die reale Welt ist dementsprechend eine neue Realität entstanden, in der jeder wahrnehmbare Moment einer einheitlich definierten Zeit, beziehungsweise einem Zustand des Uhren-Modells, entspricht. Diese Wahrnehmung von Zeit ist mittlerweile so im modernen Menschen verankert, dass ein andere Vorstellung kaum noch möglich ist (vgl. Weizenbaum 1976, 20-24).

Mit diesen Entwicklungen einher geht eine Abkehr von direkter Wahrnehmung hin zur Überprüfung eines abstrakten Modell-Zustands. Mit der Verbreitung der Uhr änderte sich die Struktur des Alltags fundamental: Anstatt sich von der Wahrnehmung des Sonnenaufgangs wecken zu lassen, begannen Menschen ihr Aufwachen nach der Anzeige der Uhr auszurichten. Die Uhr hat damit eine neue Realität geschaffen, nach der Menschen sich ausrichten müssen. Für Weizenbaum stellt diese neue Realität eine schlechtere Version der vorherigen dar, da sie den Menschen dazu zwingt, seine direkten

Sinneswahrnehmungen zu ignorieren und sich nach den Vorgaben der Uhr zu richten. So waren es nicht mehr natürliche Gefühle wie Hunger oder Müdigkeit nach denen ein Mensch gegessen und geschlafen hat, sondern feste Uhrzeiten, vorgegeben von der Uhr. Diese Entwicklung, subjektive Erfahrung zugunsten von objektiven Messungen zurücktreten zu lassen, sieht Weizenbaum als Basis für die moderne Wissenschaft. Letztendlich führte diese Entwicklung dazu, so Weizenbaum, dass nur noch Quantifizierungen als legitime Beschreibungen der Realität zulässig geworden sind: "[...] experiences of reality had to be representable as numbers in order to appear legitimate in the eyes of common wisdom" (vgl. Weizenbaum 1976, 25). Im Computer sieht Weizenbaum eine Verkörperung dieser Idee, dass vor allem quantitative und mathematische Beschreibungen als legitim betrachtet werden. Als Medium verfügt der Computer konkret über mehrere Eigenschaften, die Artefakte dieser Ideen sind. Er ist durch seine Regularität und den Folgen innerer Gesetze definiert, deren Regeln er blind folgt. Während Maschinen vormalig als etwas betrachtet wurden, das Energie manipuliert, hat sich die Aufgabe des Computers als Maschine zur Manipulation von Information verändert (vgl. Weizenbaum 1976, 41). Obwohl der Computer nach Weizenbaum lediglich ein Vehikel für diese Ideen ist, lässt sich ein selbstverstärkender Prozess aus dieser Einsicht ableiten; je mehr der Computer eingesetzt wird, insbesondere wenn er erfolgreich eingesetzt wird, desto mehr wird dadurch die Position gefestigt, dass primär oder sogar exklusiv ein quantitativer Zugang zur Welt legitim ist. Dies führt ultimativ dazu, dass durch eine verbreitete, erfolgreiche Nutzung des Computers quantitative Daten und Beschreibungen privilegiert sind gegenüber allen anderen Betrachtungsformen. Dieser Prozess erhält sich selbst, denn Wissenschaft und Technologie erhalten sich selbst durch ihre erfolgreiche Anwendung aus der Macht und Kontrolle resultieren (vgl. Weizenbaum 1976, 130).

In "Computer Power and Human Reason" wollte Weizenbaum auch erforschen, warum sich Menschen so schnell auf eine para-soziale Interaktion mit einem simplen Computerprogramm wie ELIZA einließen und sogar Fachleute wie Psychiater in dem Programm eine legitime therapeutische Unterstützung vermuteten. Für Weizenbaum gibt es mehrere Eigenschaften des Computers, die diese Reaktionen erklären können. Wie bereits beschrieben gab es auch bereits vor der Erfindung des Computers eine enge Verflechtung zwischen Menschen und ihren Werkzeugen. Seit der Entwicklung einfacher

Werkzeuge probieren Menschen, ihre natürlichen Fähigkeiten mit künstlichen Mitteln zu erweitern und spätestens seit der Industrialisierung leben Menschen in einer Welt, in der Werkzeuge und Maschinen eine Rolle in fast allen Aspekten und Bereichen des alltäglichen Lebens spielen. Um Maschinen nutzen zu können, müssen Menschen deren Funktionsweise zu einem gewissen Grad kinästhetisch und in ihrer Wahrnehmung internalisiert haben. Durch diesen Prozess der Internalisierung werden Werkzeuge ein Teil des Menschen und verändern dadurch nicht nur den Menschen, sondern auch das Verhältnis des Menschen zu sich selbst. Werkzeuge und Maschinen haben geschichtlich betrachtet primär wie Prothesen existierende menschliche Fähigkeiten erweitert und verlängert; der Computer hingegen stellt in dieser Hinsicht einen Paradigmenwechsel dar. Anstatt lediglich die Fähigkeiten des menschlichen Körpers zu verlängern, ist der digitale Computer die erste Maschine, die bestimmte intellektuelle Funktionen ausführt. Damit verändert sich nach Weizenbaum auch die Beziehung zwischen Mensch und Maschine grundlegend; eine Maschine, die die kognitiven und emotionalen Möglichkeiten des Menschen erweitert, lässt eine deutlich intensivere Verbindung mit dem Menschen zu, als ein Werkzeug wie der Hammer, der lediglich die Muskelkraft des Menschen verstärkt. Weizenbaum schlussfolgert, dass Menschen als höchst adaptive Lebewesen nicht mehr hinterfragen, wie schnell und intensiv sie sich an den Computer koppeln (vgl. Weizenbaum 1976, 5-9).

Der Computer stellt als erste “intellektuelle Maschine” damit auch einen Unterschied zu bestehenden Technologien wie der Schrift dar, die auch die menschliche Kognition erweitern. Im Gegensatz zur Schrift, die letztendlich auch als “prosthetische” Technologie betrachtet werden kann, ist der Computer eine autonome Maschine in der Tradition der Uhr. Dies bedeutet dass der Computer, obwohl natürlich von Menschen programmiert, in einem bestimmten Sinne eigene Entscheidungen trifft und Menschen nur eine begrenzte Kontrolle über ihn haben. Dies liegt laut Weizenbaum daran, dass der Computer nicht wie eine klassische mechanische Maschine mit leicht nachvollziehbaren Abläufen arbeitet. Während Laien davon ausgehen dass ein Computer lediglich exakte Befehle ausführt und deswegen sein Verhalten völlig vorhersagbar ist, sieht Weizenbaum den Computer eher wie eine Bürokratie als wie eine klassische Maschine aufgebaut. Statt simpler interner Abläufe gleichen die Prozeduren innerhalb des Computers einer Bürokratie mit vielen verschiedenen Instanzen die Daten bewerten, Entscheidungen treffen und Informationen

weiterleiten. Dadurch wird es nicht nur für die Benutzer, sondern sogar für die Entwickler eines Computerprogramms unmöglich, die genaue Entscheidungsfindung eines Computerprogramms nachzuvollziehen oder das Resultat vorherzusagen. Diese “interne Bürokratie” eines Computerprogramms entsteht insbesondere bei der Entwicklung von sehr großen Programmen, an denen eine Vielzahl von Programmierern mitgewirkt hat, ohne dass eine einzelne Person einen vollständigen Überblick über den Entwicklungsprozess hat (vgl. Weizenbaum 1976, 249-251). Dies führt paradoxerweise gleichzeitig zu einer Zentralisierung als auch Diffusion jeglicher Verantwortung. Wenn sich alle Parteien bei Entscheidungen auf den Computer beziehen, dessen Programme effektiv autoren- und damit verantwortungslos sind, erhält der Computer eine zentrale Verantwortung. Da der Computer aber eine Maschine ist, die selbst seine Entwickler nicht mehr komplett verstehen, und der nicht über die Fähigkeit menschlicher Entscheidungsfindungen verfügt, kann der Anspruch an die Verantwortung des Computers durch die Maschine nicht erfüllt werden. Da sich die Hintergründe hinter der Entscheidung des Computers nicht weiter erforschen lassen – es könnten bestimmte Werte, ethische Belange, theoretische Erwägungen, und so weiter dahinter stehen – wird die Aussage “weil der Computer es sagt” zur finalen Entscheidung. Somit ist am Ende niemand, weder Menschen noch der Computer, wirklich verantwortlich (vgl. Weizenbaum 1976, 254).

Weizenbaum kritisiert zwei Entwicklungen, die die Konsequenz dieser Eigenschaften des Computers und seines Entwicklungsprozess sind. Zum einen geben Menschen und Gesellschaften immer mehr Verantwortung an Computersysteme ab, obwohl diese Systeme immer komplexer und unverständlicher werden. Wichtige Entscheidungen werden so an Computer abgegeben, ohne dass die Kriterien und Regeln nach denen das Programm arbeitet für Benutzer oder Entwickler des Programms vollständig verständlich und nachvollziehbar sind. Durch diese Intransparenz erhalten Computerprogramme eine Form von Selbstlegitimierung. Da es keine einzelne Instanz gibt, die die internen Vorgänge des Computers beziehungsweise spezifischen Programms versteht, kann eine falsche Änderung das Programm unbrauchbar machen. Die einzige mögliche Modifikation an diesen Programmen ist damit die Erweiterung – ein intransparentes, aber funktionierendes System kann erweitert, aber nicht mehr grundsätzlich verändert werden. Dadurch, dass Menschen ein großes Vertrauen in die Entscheidungen dieser ständig

wachsenden Systeme investieren, entsteht eine konstante Selbstlegitimierung der internen Regeln dieser Computersysteme (vgl. Weizenbaum 1976, 236).

Zusätzlich zu diesen beiden Eigenschaften – der intransparenten Komplexität und dem Fehlen eines einzelnen "Autoren" – zeichnet sich der Computer dadurch aus, dass er nur Informationen und Anweisungen nur in einem bestimmten Format akzeptiert. Wie bereits erläutert müssen Prozeduren für den Computer in einer formalen Sprache kodiert sein, die dem Computer exakte, unzweideutige Befehle gibt. Außerdem müssen die Daten mit denen der Computer arbeitet in einem bestimmten Format bestehen, nämlich primär numerisch-quantitativ. Durch Nutzung des Computers werden Daten in diesem Format so gegenüber anderen Daten privilegiert und letztendlich als einzige legitime Form von Information angesehen. Weizenbaum sieht darin das Potenzial, dass der Computer ein Instrument zur Zerstörung von Geschichte wird; alle historischen Daten, die nicht in diesem Format vorliegen, werden durch den Computer als primäres Speichermedium zerstört oder vergessen (vgl. Weizenbaum 1976, 238). Auch diese Eigenschaft des Computers, nur bestimmte Formen von Daten zuzulassen, sieht Weizenbaum als Verkörperung einer reduktionistischen Wissenschaftsmethodik. Diese Methodik basiert darauf, die Realität zu simplifizieren und zu abstrahieren und alle Formen von Daten zu ignorieren, die nicht in ein bestimmtes Framework passen: "Science can proceed only by simplifying reality. The first step in its process of simplification is abstraction. And abstraction means leaving out of account all those empirical data which do not fit the particular conceptual framework within which science at the moment happens to be working [...]" (vgl. Weizenbaum 1976, 5-9).

Wegen dieser Eigenschaften beschreibt Weizenbaum diese komplexen, autorenlose und quantitativ arbeitenden Computersysteme als etwas, das mit einer vermeintlichen wissenschaftlichen Autorität Entscheidungen trifft. Diese Autorität lässt, so Weizenbaum, keine Nachfragen nach dem Wahrheit oder Gerechtigkeit dieser Entscheidungen zu: "[...] the reification of complex systems that have no authors, about which we only know that they were somehow given us by science and that they speak with its authority, permits no questions of truth or justice to be asked" (Weizenbaum 1976, 252). Diese Aussage basiert er auf einer wissenschaftskritischen Argumentation von Horkheimer; nach dieser lassen

sich Werte, Gefühlszustände, ethische Urteile und ähnliche Ideen nicht mit den Methoden der empirisch-materialistischen Wissenschaft beschreiben oder verstehen. Die Frage zum Beispiel, ob Gleichheit besser als Ungleichheit ist, hat im Kontext der empirischen Wissenschaft keine Bedeutung. Da der Computer für Weizenbaum eine Verkörperung gerade dieser "Mechanisierung von Vernunft" darstellt, sieht er die Arbeitsweise des Computers als inkompatibel mit der Klärung dieser Wertfragen an (vgl. Weizenbaum 1976, 252).

Auch auf einer gesellschaftlichen Ebene führt die Akzeptanz des Computers zu einer Diffusion von Verantwortung und missgeleitetem Vertrauen in quantitative Daten und die technische Autorität des Computers. Als frühes Beispiel für diese Konsequenzen führt Weizenbaum den Einsatz von Computern im Vietnamkrieg an. Im Vietnamkrieg wurden Computer eingesetzt, um auf Basis von maschinen-lesbaren Daten anderer Computer Koordinaten für Luftangriffe zu berechnen. So nutzten Offiziere ohne jegliche Kenntnis über die interne Funktionsweise dieser Maschinen Computer, um bestimmte Gebiete Vietnams zu sogenannten "free-fire zones" zu erklären, wenn diese eine bestimmte Dichte an Vietcong-Kämpfern erreicht hatten. Das große Vertrauen, dass die Öffentlichkeit in die vermeintliche Autorität quantitativer Daten setzt, zeigt sich laut Weizenbaum in der nicht genehmigten Bombardierung Kambodschas im Jahr 1969. Als Nixon ohne Zustimmung des Kongresses das neutrale Kambodscha bombardieren ließ, gelang dies nur indem Kongressabgeordneten und der Öffentlichkeit gefälschte Berichte vorgelegt wurden laut denen die Angriffe auf vietnamesische Ziele nahe der kambodschanischen Grenze geflogen wurden (Hersh 1973). Schlussendlich wurde diese Fälschung enttarnt, da die die Computer für die echten Angriffe immer noch exakten Daten brauchten, welche letztendlich veröffentlicht wurden. Für Weizenbaum zeigt dieses Beispiel, wie alle Parteien ihre Wahrnehmung und Verantwortung an die interne Logik des Computers abgegeben haben: Offiziere haben Zielkoordinaten bestimmt auf Basis dessen, was der Computer ihnen vorgegeben hat. Die Öffentlichkeit ließ sich durch die gefälschten Berichte täuschen, die aber durch ihre maschinelle Autorität zuerst glaubwürdig erschienen. Selbst der höchste Admiral während des Vietnamkriegs, Admiral Hoover, sagte, dass er seine Urteile auf das stützen muss, "was der Computer sagt" (New York Times 1973). Gesamtgesellschaftlich führt dies zu einer Situation, in der anonyme Kräfte ohne

Verantwortung wichtige Fragen bestimmten und gleichzeitig den Rahmen möglicher Antworten vorgeben (vgl. Weizenbaum 1976, 240).

2.3.2 Der Einsatz von Computern

An die Frage, über welche Eigenschaften der Computer als Werkzeug beziehungsweise Medium verfügt, schließt sich für Weizenbaum die Frage nach dem Einsatz des Computers an. Diese Frage betrachtet Weizenbaum insbesondere im Kontext von menschlichen Entscheidungsprozessen, die von Computern übernommen oder unterstützt werden sollen. Relevant sind hierbei nicht nur die Möglichkeiten die durch den Computer und Künstliche Intelligenz geschaffen werden, sondern insbesondere die Grenzen dieser Einsatzmöglichkeiten.

Auf den ersten Blick erscheint der Computer als eine Universalmaschine, die jedes Problem lösen kann. Der Eindruck, dass ein Computer auch technisch für wirklich jeden Zweck eingesetzt werden kann, ist ein Fehlschluss der auf einem Missverständnis des Worts “Universalmaschine” beruht. In technischer Hinsicht ist dies nicht komplett falsch; jeder Computer ist eine Turing-Maschine, die grundsätzlich jede “effective procedure” ausführen kann. Diese Prozeduren besteht aus Regeln und Abläufen, die eindeutig formuliert und in einem für den Computer nutzbaren kodierten Format verfügbar sein müssen. Nur Prozesse, Abläufe und Entscheidungen die sich in dieser Form beschreiben lassen, können auch von einem Computer verstanden und ausgeführt werden. Diese Daten müssen, wie bereits beschrieben, auf eine mathematisch-quantitative Form reduziert worden sein. Phänomene und Fragen die sich nicht in einer solchen Form ausdrücken lassen, können dementsprechend auch nicht mit Computern bearbeitet werden. Dies bedeutet, dass die Einsatzmöglichkeiten des Computers dort enden, wo die mathematischen Beschreibungen einer Problemstellungen enden. Insbesondere Im Kontext der Frage nach dem Einsatz von Computern um Entscheidungen zu treffen, die vormals Menschen getroffen haben, ergibt sich die grundsätzliche Frage, ob sich überhaupt alle menschlichen Entscheidungsprozesse in einer solchen Eindeutigkeit formulieren lassen, um sie von einem Computer verarbeiten lassen zu können: “Are all the decisionmaking processes that humans employ reducible to effective procedures and hence amenable to machine computation?” (vgl. Weizenbaum 1976, 67). Weizenbaum

akzeptiert, dass es durchaus simple menschliche Entscheidungsprozesse gibt, die sich formalisieren und verarbeiten lassen. Trotzdem wissen Menschen mehr, als sie sagen beziehungsweise beschreiben können: “We know more than we can tell” (vgl. Weizenbaum 1976, 71). Damit beschreibt er einen, für seine Theorie, bedeutenden Unterschied zwischen intuitivem menschlichem Wissen und der Möglichkeit, dieses auszudrücke. Diesen Unterschied zwischen Wissen und Ausdrücken beziehungsweise Verstehen beschreibt er am Beispiel einer Stadtkarte und dem Wissen über die Regeln von Schach. Nach Weizenbaum ist die Verfügbarkeit von formalisiertem Wissen, wie die Kenntnis einer Stadtkarte oder den Regeln eines Schachspiels, nicht gleichzusetzen, eine Stadt zu kennen oder Schach wirklich verstanden zu haben: “[...] earlier I said that to have a map of a city is not to know the city. Similarly, to be able to tell the rules of chess is not to know chess. The chess master knows more than he can tell [...]” (vgl. Weizenbaum 1976, 72). Die allgemeine Frage, ob sich alles in einer für einen Computer nutzbaren formalisierten Form ausdrücken lässt, beantwortet Weizenbaum deutlich negativ: “‘Can anything we may wish to do be described in terms of an effective procedure?” The answer to that question is ‘No’.” (vgl. Weizenbaum 1976, 65). Sprache ist ein wichtiges Element in diesem Unterschied zwischen dem Menschen und dem Computer. Diese Prozesse lassen sich nur mit Sprache beschreiben. “Effective procedures” lassen sich nur in einer formalen Sprache beschreiben, die eine eindeutige Beschreibung dieser Prozeduren möglich macht. Natürliche Sprachen, wie sie von Menschen benutzt werden, sind nicht rein formal. Auch natürliche Sprachen haben formale Aspekte, der signifikante Unterschied zu rein formalen Sprachen ist, dass sie eine Bedeutung in sich tragen. Weizenbaum positioniert sich bei dieser Frage auch gegenüber anderen Wissenschaftlern seiner Zeit, die diesen Unterschied zwischen Wissen und Verstehen nicht machen. So setzt Marvin Minsky das Verstehen von Musik und Kunst gleich mit der Fähigkeit, Computerprogramme schreiben zu können die Musik und Kunst erzeugen können (vgl. Weizenbaum 1976, 248). Weizenbaum kritisiert an dieser Position, dass solche Computer unser Verständnis von diesen Dingen nicht erweitern können. Insbesondere die bereits beschriebenen autorenlosen Programme, die selbst von ihren Entwicklern nicht mehr verstanden werden und deren innere Operationen nicht observiert werden können, können niemals das menschliche Verständnis von beispielsweise Kunst erweitern (vgl. Weizenbaum 1976, 235). So bleibt auch das vermeintliche Verständnis eines Computers immer oberflächlich und lediglich formalisiertes Wissen. Diese Unterscheidung zwischen Wissen und

Verstehen erhält eine ethische Dimension, wenn es nicht um die Regeln eines Spiels geht sondern um menschliche Belange. Da Computer nur über formalisiertes Wissen ohne Bedeutung verfügen können, muss jede künstliche Intelligenz zwangsweise völlig fremd gegenüber menschlichen Anliegen sein. Die Frage, ob man einem Computer beibringen kann was ein Richter oder Psychiaterin weiß, beides Rollen für die intuitives Wissen außerordentlich wichtig ist und die Entscheidungen mit großen Effekten treffen, hält Weizenbaum für monströs und als Zeichen des Wahnsinns seiner Zeit (vgl. Weizenbaum 1976, 226).

Doch auch wenn ein Computer ein bestimmtes Problem lösen kann, stellt dies noch keine legitime Anwendung dar. Weizenbaum betont am Beispiel der Astrologie, dass es keinen Zusammenhang gibt zwischen der Validität einer Methode und wie schnell diese ausgeführt werden kann und dies nicht verwechselt werden sollte: “If astrology is nonsense, then computerized astrology is just as surely nonsense” (vgl. Weizenbaum 1976, 35). Die pure Schnelligkeit und analytische Rechenkapazität des Computers lässt also andere wichtige Aspekte in den Hintergrund treten – Menschen sind so überwältigt von den Möglichkeiten des Computers dass sie vergessen, dass diese Leistung zielgerichtet und methodisch unterbaut sein muss. Die grundsätzliche Frage nach den Fähigkeiten eines Computers sollte sich deswegen also nicht auf dessen Schnelligkeit und Rechenkapazität beziehen, sondern auf die Möglichkeiten der Formalisierung. Nur wenn sich ein Sachverhalt, eine Prozedur oder ein Entscheidungsprozess in einer für den Computer verarbeitbaren Form formalisieren lässt, kann ein Computer dafür genutzt werden.

Im Prozess der Formalisierung für den Computer, beziehungsweise Modellbildung, existieren weitere Herausforderungen und mögliche Probleme. Modellbildung ist, so Weizenbaum, immer ein Prozess der Reduktion bei dem wichtige Entscheidungen getroffen werden, welche Aspekte eines Modells relevant sind und welche nicht. Diese Entscheidungen werden von Menschen getroffen und haben damit auch immer einen politischen und kulturellen Aspekt, den das Modell widerspiegeln wird: “We select, for inclusion in our model, those features of reality that we consider to be essential to our purpose. In complex situations [...], the very act of choosing what is essential and what is

not must be at least in part an act of judgement, often political and cultural judgement. And that act must then necessarily be based on the modeler's intuitive mental model. But again, judgement must be exercised to decide what the something might be, and whether it is 'essential' for the purpose the model is intended to serve. The ultimate criteria, being based on intentions and purposes as they must be, are finally determined by the individual, that is human, modeler" (vgl. Weizenbaum 1976, 149). Doch diese Erkenntnis, dass der vermeintlich objektive Computer mit Modellen arbeitet die implizit oder explizit menschliche Werturteile beinhalten und reproduzieren, wird innerhalb der Wissenschaftsgemeinde nicht akzeptiert oder sogar aktiv ignoriert. Laut Weizenbaum werden Werturteile mit ethischen und gesellschaftlichen Implikationen zu "rationalen" Entscheidungen umgedeutet, die als frei von jeglichen ethischen Fragen entstanden sind (vgl. Weizenbaum 1976, 263). In diesem Kontext unterscheidet Weizenbaum auch zwischen dem Entscheiden ("deciding") und Wählen ("choosing") um den Unterschied zwischen maschinellem und menschlichem Handlungsvermögen deutlich zu machen: "Instrumental reason can make decisions but there is all the difference between deciding and choosing" (vgl. Weizenbaum 1976, 259). Die bereits erwähnten objektiven Entscheidungen des Computers sind damit ein Akt des maschinellen Entscheidens zwischen mehreren Optionen, aber nicht des Wählens. Maschinelles Entscheiden basiert auf programmierten Routinen. Die Begründung einer maschinellen Entscheidung kann sich damit lediglich auf diese Programmierung beziehen. Eine menschliche Entscheidung, beziehungsweise Wahl, begründet sich aus der freien Entscheidung eines Menschen (vgl. Weizenbaum 1976, 260). Aus dem Grund schlussfolgert Weizenbaum auch, dass Computer keine gesellschaftlich wichtigen Entscheidungen übernehmen sollten. Computer können bestimmte Entscheidungen treffen und diese Entscheidungen können auch richtig sein. Die relevante Frage ist aber nicht, ob sie es können oder ob es technisch möglich ist, sondern ob Computer diese Entscheidungen überhaupt treffen sollten. Da wir Computer nicht "weise" machen können, sollten wir ihnen keine Aufgaben abverlangen, die Weisheit benötigen (vgl. Weizenbaum 1976, 226).

Bei der Frage nach den Unterschieden zwischen Mensch und Maschine stellt sich die Frage nach einer Essentialität, beziehungsweise was den Kern von etwas darstellt. Der Prozess der Wissenschaft verlangt, dass Phänomene und Objekte aus einer bestimmten Perspektive betrachtet werden müssen, die grundsätzlich reduktionistisch ist. Erst durch

den Fokus auf einige wenige Aspekte können diese betrachtet werden, auch wenn dies bedeutet dass andere Aspekte und eine ganzheitliche Betrachtung verschwinden. Während dieses Prozesses begehen nach Weizenbaum viele Menschen den Fehler, das Modell eines Objekts mit dem Objekt an sich zu verwechseln. Dies führt zu dem Glauben, dass das Objekt die gleichen Eigenschaften hat wie das Modell. Weizenbaum führt das Beispiel an, das menschliche Gehirn als mathematisches Modell zu beschreiben: “Such a model would, of course, be in principle describable in strictly mathematical terms. This might lead some people to believe that the language our nervous system uses must be the language of our mathematics. Such a belief would be an error of the kind we mean” (vgl. Weizenbaum 1976, 150). Aus diesem Missverständnis heraus denken selbst Akademiker dass der Menschen “nicht mehr” als eine reine Datenverarbeitungsmaschine ähnlich dem Computer sei. So zitiert Weizenbaum den Begründer der Kognitionspsychologie George Miller: “Many psychologists have come to take for granted in recent years [...] that men and computers are merely two different species of a more abstract genus called ‘information processing systems’” (vgl. Weizenbaum 1976, 158). Diese Abstrahierung des Menschen geht sogar so weit, dass Weizenbaums Zeitgenossen aus der KI-Forschung selbst Psychiater als regelbasierte Maschinen betrachten, die festen Regeln folgen um Patienten systematisch und zielgerichtet zu heilen. Weizenbaum hält dies für ein massives Problem – ein Psychiater, der seinen Patienten nicht empathisch als menschliches Gegenüber sondern als technisch lösbares Problem betrachtet, missversteht die Natur seiner Arbeit (vgl. Weizenbaum 1976, 21).

Nach Weizenbaum gibt es jedoch signifikante Unterschiede zwischen Menschen und Maschinen. Ein wichtiger Aspekt spielt dabei die Nutzung von Sprache. Sprache ist eng verwoben mit der menschlichen Fähigkeit zu Denken und ein Computer müsste diese Form von Sprache verstehen, um mit dem Menschen vergleichbar zu sein: “Man’s capacity to manipulate symbols, his very ability to think, is inextricably interwoven with his linguistic abilities. Any re-creation of man in the form of machine must therefore capture this most essential of his identifying characteristics.” (vgl. Weizenbaum 1976, 184). Neben ihrer sprachlichen Fähigkeiten sind Menschen definiert durch die Herausforderungen, denen sie in ihrer evolutionären Geschichte begegnet sind. Diese Herausforderungen entstammen den biologischen Bedürfnissen des Menschen und sind somit unmöglich von

Computern, die diese Bedürfnisse nicht teilen, zu verstehen oder zu simulieren (vgl. Weizenbaum 1976, 223).

2.3.3 Zusammenfassung

Weizenbaum wollte in “Computer Power and Human Reason” nach Gründen für die überraschenden Reaktionen auf ELIZA suchen; die Schnelligkeit mit der Menschen eine soziale Beziehung zu ELIZA aufbauten, die Bereitschaft von Therapeuten ihre Arbeit an ELIZA abzugeben und die Begeisterung der akademischen Gemeinde über das vermeintlich gelöste Problem der Sprachverarbeitung. Den Ursprung dieser Reaktionen sieht er in einem bereits existierenden Weltbild, in dem die Welt “zu einem Computer gemacht” wurde, also durch eine reduktionistische Perspektive betrachtet wird. Dieses Weltbild hat sich mit dem Fortschritt der modernen Wissenschaftsphilosophie entwickelt. Es ist geprägt durch den Vorrang von einem quantitativen Zugang zur Welt gegenüber persönlicher Erfahrung und einem starken Fokus auf formale Modelle, die die vermeintlich wichtigsten Charakteristiken eines Phänomens beschreiben.

Obwohl der Computer für Weizenbaum lediglich eine Verkörperung dieser bereits bestehenden Ideen darstellt, stellt er doch einen Paradigmenwechsel in der Beziehung zwischen Mensch und Maschine dar. Weizenbaum stellt den Computer in die Tradition der Uhr, die für ihn die erste relevante autonome Maschine darstellt. Durch die Einteilung von Zeit in diskrete Einheiten hat die Einführung der Uhr eine neue Realität geschaffen in der sich Menschen nach dem internen Modell der Uhr richten müssen. So wie die Modellierung von Zeit in Form der Uhr eine neue Realität geschaffen hat, schafft auch die Modellierung des Menschen als biologische Maschine eine neue Realität. Der Computer ist das erste Werkzeug, dass eine intellektuelle Erweiterung des Menschen darstellt. Damit kann der Mensch zum ersten Mal in seiner Geschichte kognitive Aufgaben an eine Maschine übergeben. So müssen sich aber auch kognitive Prozesse und Vorgänge, die vormals den Limitierungen des Menschen unterlagen, nun den Eigenschaften des Mediums Computer anpassen. Während Menschen Informationen und Sinneseindrücke jeglicher Art wahrnehmen und verarbeiten können, müssen diese Daten für den Computer in einer quantitativen Form formalisiert sein, was zu einer Privilegierung dieses Formats gegenüber anderen Daten führt. Die inneren Prozesse eines Computers können

außerdem so komplex werden, dass selbst die Entwickler eines Programms die Resultate nicht mehr vollständig nachvollziehen können. Dadurch entsteht die zirkuläre Begründung von Computer-Entscheidungen – “weil der Computer es sagt” –, die zu einer Diffusion oder sogar Verschwinden von Verantwortung für die Resultate des Computers führt. Diese Komplexität führt außerdem zu einer Selbstlegitimierung, da die Programme nicht mehr verändert sondern lediglich erweitert werden können.

Da ein Werkzeug beziehungsweise Maschine nun erstmals kognitive Aufgaben übernehmen kann, kann auch erstmals eine Maschine Entscheidungen in menschlichen Belangen treffen. Dieser Idee widerspricht Weizenbaum und argumentiert, dass es signifikante Unterschiede zwischen dem Menschen und Computer gibt, weswegen Computer für bestimmte Zwecke keine Entscheidungen treffen sollten. So können Computer zwar in formalisierter Form über Regeln und Prozeduren ein Wissen über bestimmte Dinge aufbauen, dies ist aber grundsätzlich unterschiedlich zu dem intuitiven Wissen das Menschen über ihre Umwelt haben. Computer können dementsprechend auch keine Werturteile treffen, denn Werturteile lassen sich nicht quantitativ operationalisieren – die Frage, ob beispielsweise Freiheit mehr wiegt als Sicherheit, lässt keine quantitativ-mathematische Berechnung und Antwort zu. Da die Form der Entscheidungsfindung des Computers deswegen grundlegend unterschiedlich ist von der Art wie Menschen zu Entscheidungen kommen, sollten Computer keine Entscheidungen in menschlichen Belangen treffen. Weizenbaum stellt damit klar, dass die entscheidende Frage nicht ist, ob Computer technisch betrachtet bestimmte Entscheidungen treffen können, sondern ob sie es sollten. Die zentrale Frage ist damit keine technische, sondern eine ethische Frage.

Nicht alle von Weizenbaum in “Computer Power and Human Reason” erwähnten Themen können oder sollen in dieser Arbeit diskutiert werden. Im Folgenden sollen deswegen zwei Themen mit der größten medienwissenschaftlichen Relevanz intensiver behandelt werden. Zuerst soll Weizenbaums These vom Computer als Werkzeug, beziehungsweise Metapher, und dessen gesellschaftlichen Einfluss im medientheoretischen von Marshall McLuhan analysiert werden. Für Weizenbaum hat der Computer als Medium das gesellschaftliche Denken bereits soweit verändert, dass die weit verbreitete Annahme

besteht, dass Computer für alle Rollen und Funktionen eingesetzt werden können. Weizenbaum hält diese Annahme für fehlgeleitet und beschreibt Funktionen, die nicht von Computern übernommen werden sollten – um Weizenbaums theoretischen Ausführungen ein praktisches Gewicht zu verleihen, sollen seine Ideen zu den Grenzen des Computers mit Beispielen aus der modernen algorithmischen Entscheidungsfindung (ADM) kontextualisiert werden. Während diese Systeme nämlich mittlerweile zum Alltag gehören, schrieb Weizenbaum in den 70ern noch davon, dass es bisher keine breiten Anwendungen aus der KI-Forschung gab (vgl. Weizenbaum 1976, 244).

2.4 Medientheoretischer Kontext

Obwohl sich Weizenbaum in “Computer Power and Human Reason” nicht explizit auf andere Medientheorien bezieht und von der Computer-”Metapher”, nicht vom Computer-Medium, spricht, sind seine Überlegungen zum Computer und dessen gesellschaftlichen Einfluss deutlich medientheoretisch beeinflusst. Dies wird am deutlichsten, wenn Weizenbaums Gedanken im Kontext von Marshall McLuhans Medientheorie betrachtet werden, die in den 1960er-Jahren mit seinen Werken populär wurde.

Marshall McLuhan leitete als bekannter Theoretiker der damaligen Zeit mit seinen Medientheorien einen Paradigmenwechsel ein; anstatt sich auf den Inhalt eines Mediums zu fokussieren, legte McLuhan seinen Forschungsschwerpunkt auf das Medium selbst und den Einfluss, den Medien auf Gesellschaften haben und formulierte diese Idee als “the medium is the message”. Der Ansatz, gesellschaftliche und soziale Konfigurationen als Resultat der primär genutzten Medien zu betrachten, war nicht vollkommen neu sondern inspiriert von den Ideen Harold Innis’. Sowohl Innis als auch McLuhan sahen Medien nicht einfach nur als technologische Begleiterscheinungen einer Gesellschaft, sondern als zentrale Determinanten des gesellschaftlichen Zusammenlebens (Carey 1967). Innis war insbesondere daran interessiert, welche Rolle Kommunikationsmedien in der Entwicklung von Zivilisationen spielen und auf welche Weise unterschiedliche Kommunikationsmedien zu unterschiedlichen Kulturpraktiken und Regierungsformen führten (Babe 2000). Ausgehend von Innis’ Ansatz, kausale Zusammenhänge zwischen

staatlicher Organisation und den primären Medien einer Gesellschaft herzustellen, entwickelte McLuhan die These von Kausalzusammenhängen zwischen einer Gesellschaft und ihrem Hauptmedium weiter.

Die zentrale Aussage seines Buchs “Understanding Media” ist “the medium is the message”. Damit beschreibt McLuhan seine Position, dass der eigentliche Inhalt eines Mediums irrelevant ist gegenüber den Qualitäten des Mediums an sich. McLuhan nutzt für diese Analyse eine sehr breite Definition des Begriffs Medium; für ihn sind Medien alles, was die menschlichen Fähigkeiten erweitert (vgl. McLuhan 1994, 7). Dazu gehören Kommunikationsmedien und andere Technologien, neue Verfahren die mit den technologischen Entwicklungen einhergehen aber auch kulturelle Praktiken. Unter die großen technologischen Entwicklungen fallen für McLuhan beispielsweise auch die Eisenbahn und Luftfahrt (vgl. McLuhan 1994, 8), Automatisierung oder das Geldsystem (vgl. McLuhan 1994, 18). McLuhan stellt die These auf, dass sich die sozialen Effekte eines Mediums nicht über die vermittelten Inhalte erforschen lassen. Die einzige relevante Botschaft (“message”) eines Mediums ist nach McLuhan die Veränderung in Größe, Geschwindigkeit oder Methode, die das Medium im menschlichen Zusammenleben auslöst: “For the "message" of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs” (vgl. McLuhan 1994, 8).

Nach McLuhan sind Gesellschaften von ihren Medien in vergleichbarer Weise abhängig wie von ihren natürlichen Rohstoffen; so wie die Verfügbarkeit bestimmter Rohstoffe Kulturen prägt, prägen auch die genutzten Medien Kulturen und ihr Zusammenleben (vgl. McLuhan 1994, 21). Ein Beispiel für diese gesellschaftliche Veränderung ist, was McLuhan mit dem Begriff des “Global Village” ausgedrückt hat (McLuhan 1994, 93) Während Menschen vor der Entwicklung elektronischer Kommunikationsmedien lange Zeit in relativ kleinen Gemeinschaften lebten und primär innerhalb dieser Gemeinschaften kommunizierten, führte die Einführung von Technologien wie dem Telegraphen zu einer Erweiterung der Kommunikationsmöglichkeit des Einzelnen und damit gleichzeitig zu einer Verkleinerung der Welt. Da im Prinzip nun alle anderen Menschen auf der Welt erreichbar waren, wie es vormals nur Menschen aus dem direkten Umfeld waren, bekam die globale Kommunikation den Charakter der Kommunikation wie in einer kleinen Gemeinschaft. So

führte die Entwicklung dieser Kommunikationsmedien zu einer Veränderung des globalen gesellschaftlichen Zusammenlebens. Medien sind dabei nicht neutral in dem Sinne, dass ihr Nutzen oder Schaden erst durch die Art des Einsatzes bestimmt wird. McLuhan erklärt dies am Beispiel der Automatisierung: es ist irrelevant ob Maschinen am Fließband Cornflakes oder Cadillacs (also ihren “Inhalt”) produzieren, da es die Methode der Automatisierung selbst ist, die das gesellschaftliche Leben verändert. Um das gesellschaftliche Zusammenleben zu verändern, braucht es auch nicht die Zustimmung oder Ablehnung der Menschen, die in der Gesellschaft leben. McLuhan macht dies an dem Beispiel von Japan im 17. Jahrhundert fest, als das Geldsystem langsam das feudale Regierungssystem abschaffte und Japan für den internationalen Handel öffnete und so die Gesellschaft in ihren Grundfesten verändert hat (vgl. McLuhan 1994, 19).

Neue Medien entstehen entstehen in einer Art von selbstverstärkendem Feedback-Prozess der auf “Selbst-Amputation” beruht. Neue Medien führen zu neuen Belastungen, beispielsweise dadurch dass sie Kommunikation schneller und weitreichender machen. Auf diese neuen Belastungen reagieren Menschen mit wieder neuen Erfindungen, die ihnen dabei helfen sollen mit den Belastungen umzugehen. Der Prozess neuer Erfindungen wird von McLuhan als Auto-Amputation beschrieben; der mit neuen Irritationen konfrontierte Körper amputiert beziehungsweise externalisiert daraufhin Funktionen in Form von externen Medien: “Any extension of ourselves they regard as "autoamputation," and they find that the autoamputative power or strategy is resorted to by the body when the perceptual power cannot locate or avoid the cause of irritation.” (vgl. McLuhan 1994, 42-42). Dieser Prozess der konstanten Weiterentwicklung und Selbstamputation ist ab einem bestimmten Punkt, der “breaking boundary”, nicht mehr aufzuhalten oder rückgängig zu machen. Ab diesem Punkt führen die Auswirkungen des Mediums zu grundlegenden Veränderungen innerhalb der Umgebung, in der es eingesetzt wird: "break boundary at which the system suddenly changes into another or passes some point of no return in its dynamic processes" (vgl. McLuhan 1994, 38).

Weizenbaums Theorie zum Computer als Werkzeug besitzt einige Parallelen zu den Theorien von Innis und McLuhan, insbesondere die zentralen Thesen zur Beziehung zwischen Medien und ihren Gesellschaften. Während Innis seinen Fokus auf die

Beziehung von Medien zu Institutionen von Herrschaft und Macht legte, bestand McLuhans primäres Interesse an dem Einfluss den Medien auf die Wahrnehmung und das Denken innerhalb einer Gesellschaft haben (vgl. Carey 1967). Weizenbaum sieht den Einfluss des Computers beiderseitig; der Computer beeinflusst als kognitive Maschine das Denken des Menschen und hat damit einen bedeutenden Einfluss auf Herrschaftsverhältnisse innerhalb einer Gesellschaft. So wie McLuhan sieht auch Weizenbaum den Treiber von technologischem Fortschritt auf Seiten der Werkzeuge, und nicht als einen Prozess der vom Menschen ausgeht. Dementsprechend hält er auch die Frage, ob der wachsende Einfluss des Computers gewollt ist, für nicht sinnvoll: “There is not the slightest hint of a question as to whether we want this future. It is simply coming.” (vgl. Weizenbaum 1976, 257). So wie Innis Medien einen maßgeblichen Einfluss auf die Herrschaftsform zuschreibt, betrachtet Weizenbaum den Computer als Werkzeug zur Aufrechterhaltung des politischen “Status Quo”. Dadurch, dass der Computer Probleme vermeintlich durch schiere Rechenleistung löst, entsteht die Tendenz, dass sich die Möglichkeit an Lösungsansätzen mehr und mehr auf den Computer beschränkt. Diese Tendenz sieht Weizenbaum auch in anderen Aspekten des gesellschaftlichen Lebens und weist darauf hin, wie grundlegend politische Probleme zu technischen Problemen umgedeutet werden, die systematisch gelöst werden können solange die richtigen Maßnahmen ergriffen werden. Als Beispiel dienen dafür die Antikriegsproteste an amerikanischen Universitäten, die von der Universitätsverwaltung als ein Kommunikationsproblem behandelt werden, das durch bessere Kommunikation gelöst werden kann und nicht als das Resultat von grundsätzlichen politischen Unterschieden (vgl. Weizenbaum 1976, 266). In Referenz an McLuhan lässt sich argumentieren, dass Weizenbaum erkannt hat, wie die Botschaft des Mediums Computers, nämlich die absolute Lösbarkeit eines jeden Problems durch die quantitative Power des Computers, sie blind gemacht hat für den Charakter des Mediums selbst: “Indeed, it is only too typical that the "content" of any medium blinds us to the character of the medium” (vgl. McLuhan 1994, 9).

Eine besondere Rolle nehmen für Weizenbaum auch als Innis die Menschen ein, die mit dem Hauptmedium ihrer Zeit umgehen können. Innis beschreibt, wie die Schreiber des alten Ägyptens durch ihre Fähigkeiten zu einer privilegierten Klasse aufstiegen: “literacy

was valued as a stepping-stone to prosperity and social rank. Scribes became a restricted class and writing a privileged profession" (Innis 2007, 37).

Trotz der Gemeinsamkeiten zwischen den medientheoretischen Ideen von Weizenbaum und McLuhan gibt es einige signifikante Unterschiede. McLuhan betrachtet alle Medien und Technologien als Erweiterungen des Menschen, die bestehende Sinne oder körperliche Funktionen erweitern und intensivieren. In diese Gruppe schließt er auch den Computer ein: "Because all media, from the phonetic alphabet to the computer, are extensions of man that cause deep and lasting changes in him and transform his environment" (Playboy Magazine 1969). Auch McLuhan betrachtet den Computer als einzigartig in der Mediengeschichte und beschreibt ihn als Erweiterung des menschlichen Nervensystems: „The computer is by all odds the most extraordinary of all the technological clothing ever devised by man, since it is the extension of our central nervous system" (McLuhan & Fiore 1968, 35). Demgegenüber steht Weizenbaums Einteilung von Medien in "prosthetische" und "autonome" Medien, beziehungsweise Werkzeuge. Während die "prosthetischen" Werkzeuge weitestgehend mit der Definition von McLuhan übereinstimmen, betrachtet Weizenbaum den Computer als autonome Maschine. Diese Unterscheidung ist signifikant, da sie eine grundsätzlich unterschiedliche Beziehung zwischen Mensch und Werkzeug offenbart. Ein den Menschen erweiterndes Medium folgt im Kern immer noch den Gesetzen der menschlichen Fähigkeit, die es erweitert. Eine autonome Maschine hingegen, wie Weizenbaum am Beispiel der Uhr beschreibt, folgt ihrer eigenen internen Logik der sich Menschen anpassen müssen. Den Ursprung dieser inneren Logik sieht Weizenbaum allerdings außerhalb der Maschine selbst; für ihn ist diese Maschinenlogik die Verkörperung einer bereits bestehenden Idee oder Gesetzes. So betrachtet er als Essenz der Maschine ihre unnachgiebige Regularität und das blinde Gehorsam den Regeln gegenüber, deren Verköperung sie ist: "[...] relentless regularity, its blind obedience of the law of which it is an embodiment.". Mit den spezifischen inneren Logiken einer autonomen Maschine befasst McLuhan sich wenig und betrachtet Medien primär im Kontext von neuen Geschwindigkeiten, die sie in einer Gesellschaft auslösen.

Die Art und Weise wie Ideologie sich in Medien als Artefakten ausdrückt ist eines von Weizenbaums zentralen Themen. Für Weizenbaum stellt der Computer die Verkörperung

eines reduktionistischen Weltbilds dar, dass den Menschen und die Welt als reduzierbar auf quantitative Modelle betrachtet. Der Computer ist deswegen so erfolgreich, weil er dieses existierende Weltbild praktisch umsetzen kann. Auch McLuhan sieht signifikante kulturelle Einflüsse in der Selektion von Medien, beschreibt diese aber nicht im Kontext von Ideologien. Ein Beispiel für diese unterschiedlichen Betrachtungsweisen sind die unterschiedlichen Perspektiven auf den IQ-Test. Weizenbaum sah den IQ-Test, beziehungsweise die Idee eines allgemeinen Intelligenzquotienten, als missgeleitet und schädlich an. Die weiträumig eingesetzten Tests führen zu dem Eindruck, dass es einen quantitativen Weg gäbe, eine feststehende und kulturell unabhängige Form von Intelligenz objektiv festzustellen. Weizenbaum stellt diesem Konzept gegenüber, dass auch Intelligenz immer abhängig von Kultur und Situation ist und es eine missgeleitete Idee ist, dass es eine Art absolute Intelligenz gäbe (vgl. Weizenbaum 1976, 203-204). Der IQ-Test ist damit auch ein Artefakt der instrumentellen Vernunft und des maschinellen Weltbilds. Auch McLuhan sah den IQ-Test als missgeleitetes Vorhaben. Während Weizenbaum die Ursache in der Tendenz zur Quantifizierung sieht, hält McLuhan die implizite kulturelle Neigung zum Medium “Typographie” als verantwortlich für den IQ-Test. Diese kulturelle Neigung zur regelmäßigen Typographie führt dazu, dass lediglich eine bestimmte Form von Verhalten, nämlich uniform, stetig und vorhersagbar, als intelligent wahrgenommen wird und stellt die Grundlage für die Idee des IQ-Tests dar: “It is in our I.Q. testing that we have produced the greatest flood of misbegottten standards. Unaware of our typographic cultural bias, our testers assume that uniform and continuous habits are a sign of intelligence, thus eliminating the ear man and the tactile man” (vgl. McLuhan 1994, 17). Weizenbaum sieht also einen Prozess, bei dem eine bestimmte Ideologie oder Weltanschauung in einem Artefakt wie dem IQ-Test umgesetzt wird, während McLuhan den IQ-Test als Ausdruck des Mediums Typographie betrachtet.

2.5 Zusammenfassung

Das Ziel dieser Arbeit ist zu überprüfen, inwiefern die Kritik von Weizenbaum aus dem Jahr 1976 im beginnenden 21. Jahrhundert noch Bestand hat. In diesem Kapitel wurden die relevanten Punkte Weizenbaums dargestellt und mit den Medientheorien von

McLuhan und Innis verglichen. Im folgenden sollen die wichtigsten Aussagen und Punkte zusammengefasst werden.

Eine zentrale Erkenntnis aus dem Vergleich mit McLuhans Medientheorie ist, dass es die Eigenschaften des Mediums Computer sind, die der Fokus der Analyse sein sollten. McLuhans Aussage “the medium is the message” folgend sind es nicht die spezifischen Algorithmen und Anwendungen des Computer, sondern der Computer selbst der analysiert werden sollte. Als Hauptmedium dieser Zeit hat der Computer selbst einen fundamental strukturierenden Effekten auf das Funktionieren moderner Gesellschaften. Die Form des Funktionierens wird dabei durch die Eigenschaften des Computers vorgegeben. Dazu gehört die Privilegierung quantitativer Daten und Modelle bei der Betrachtung von sozialen Phänomen und die Entwicklung autorenloser Programme, deren Entscheidungen sich nicht erklären oder überprüfen lassen. Dies führt zu einer Diffusion von Verantwortung, die aber gesellschaftlich akzeptiert wird da der Computer mit der mathematischen Autorität der wissenschaftlichen Methode auftritt und so seine Entscheidungen vertrauenswürdig erscheinen lässt. Der Computer verspricht durch seine enorme Rechenkapazität die absolute Lösbarkeit eines jeden Problems, unabhängig davon, ob es sinnvoll ist das Problem mit einem Computer zu lösen oder nicht.

Nach Weizenbaum ist die Akzeptanz des Computers und seinen Berechnungen das Resultat eines Weltbilds, in dem die Welt und der Mensch bereits wie eine Maschine betrachtet werden. Die Entwicklung der wissenschaftlichen Methode und in den empirischen Wissenschaften haben, so Weizenbaum, dazu geführt, dass auch der Mensch als quantifizierbares Wesen wahrgenommen wird. Daraus ergibt sich eine weitere zentrale Frage im Werk von Joseph Weizenbaum nach den Grenzen des Computers. Die Grenzen des Computers und der künstlichen Intelligenz werden durch die Operationalisierbarkeit ihrer Probleme definiert. Nach Weizenbaum lassen sich viele menschliche Belange aber nicht so operationalisieren beziehungsweise reduzieren, da diese dadurch wichtige andere Aspekte verlieren. Damit sind für Weizenbaum sind die wichtigsten Probleme zu den Grenzen des Einsatzes von Computern nicht technologisch oder mathematisch, sondern ethisch. Um diese Probleme zu lösen sollten keine Fragen nach der Machbarkeit gestellt werden (können Computer bestimmte Funktionen

übernehmen?) sondern es muss die Frage gestellt werden, ob Computer bestimmte Funktionen überhaupt übernehmen sollten (vgl. Weizenbaum 1976, 227). Die Verbreitung des Computers stellt für Weizenbaum damit die Gefahr da, dass menschliche Angelegenheiten nicht mehr nach menschlichen Maßstäben beurteilt werden, sondern nach den Maßstäben von Maschinen, die sich grundlegend vom Menschen unterscheiden (Bassett, 2019, 5).

Dieses Kapitel hat dargestellt, dass die Eigenschaften des Mediums Computer zentraler Fokus einer Analyse sein sollen, da diese soziale und ethische Folgen verursachen. In Erweiterung des Mediums Computer soll im folgenden Kapitel auf Künstliche Intelligenz und ihre Eigenschaften im speziellen eingegangen werden und die historische Entwicklung bis zum heutigen Punkt beschrieben werden.

3 Künstliche Intelligenz

3.1 Was ist Künstliche Intelligenz?

Um die Frage nach der Relevanz von Weizenbaums Positionen in einem aktuellen Kontext positionieren zu können, muss erst geklärt was Künstliche Intelligenz überhaupt ist und wie sich KI-Technologien seit ihren Anfangszeiten entwickelt haben. Zu diesem Zweck soll zuerst eine für diese Arbeit sinnvolle Arbeitsdefinition von künstlicher Intelligenz entwickelt werden.

Einer der ersten Ansätze maschinelle Intelligenz zu definieren, war das 1950 von Alan Turing entwickelte “Imitation Game”, welches später im allgemeinen Gebrauch zum “Turing-Test” wurde. Turing war interessiert an der Lösung einer fundemantentalen Frage: können Maschinen denken? Doch Turing fand diese Frage nicht passend; um sie beantworten zu können, müssten zuerst die uneindeutigen Begriffe “Maschine” und “Denken” definiert werden. Die Definition dieser Begriffe würde sich nach dem üblichen Gebrauch dieser Begriffe richten, wodurch sich ihre Bedeutung über eine Umfrage definieren ließe. Da dies absurd wäre, ersetzt Turing die Eingangsfrage mit dem Imitation Game und Fragen zu diesem. Im Imitation Game darf eine Person C einem ihm

unsichtbaren Paar, bestehend aus Mann A und Frau B, Fragen stellen und muss herausfinden, welcher der Antwortgeber eine Frau oder Mann ist. Turing ändert das Szenario nun so ab, dass einer der Antwortgeber mit einer Maschine (einem digitalen Computer bzw. einer “Universalmaschine”) ersetzt wird um die Frage zu stellen: Wird der Fragesteller genau so oft falsch liegen mit der Maschine als Antwortgeber, wie wenn beide Antwortgeber Menschen sind? Für Turing war die Frage, ob Maschinen denken können “nutzlos”; für ihn war interessant, ob eine Maschine so intelligent wirken kann, dass ihr Handeln nicht von dem eines Menschen zu unterscheiden ist (Turing 1950). Offensichtlich inspiriert von den Ideen Turings wurde 1956 während des “Dartmouth Summer Research Project on Artificial Intelligence”, das als Grundsteinlegung für KI als Forschungsbereich gilt, eine weitere Definition von künstlicher Intelligenz aufgestellt: “For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving” (vgl. Moor 2006). So wie bereits bei Turings Imitation Game wird künstliche Intelligenz daran definiert, ob sie im menschlichen Sinne intelligent ist, sondern ob sie intelligent scheint, beziehungsweise in ihrer äußeren Verhalten her nicht von menschlicher Intelligenz zu unterscheiden ist. Eine Technologie kann also dann intelligent werden, wenn sie erfolgreich die intellektuellen Fähigkeiten eines Menschen imitiert.

Die Position, maschinelle und menschliche Intelligenz anhand ihrer äußeren Erscheinung gleichzusetzen ist in den Jahrzehnten nach Turings Veröffentlichung häufiger Kritik ausgesetzt gewesen. Eine der bekanntesten Kritiken hat der Philosoph John Searle in seinem “Chinese Room”-Gedankenexperiment formuliert: ein Mensch wird zusammen mit einem Wörterbuch in einen Raum gesperrt und muss chinesische Texte übersetzen, die ihm zugeschoben werden. Durch diese Gedankenexperiment will Searle auf den Unterschied zwischen symbolischer Transformation und semantischem Verständnis aufmerksam machen; so wie der eingeschlossene Mensch lediglich nach festen Regeln chinesische Schriftzeichen im Wörterbuch findet und ersetzt ohne wirklich Chinesisch zu verstehen, hat auch ein Computer kein reales semantisches Verständnis seiner Operationen (Searle 1980). Obwohl diese philosophischen Einwände an der KI-Definition Turings und der Dartmouth-Konferenz berechtigt sind, soll diese für diese Arbeit als Arbeitsdefinition genutzt werden um KI-Technologien zu beschreiben. Sie bietet sich aus mehreren Gründen an: Diese Definition macht eine Erforschung des Themenfeldes KI

unabhängig von den spezifischen technischen Implementierungen möglich. Oft ist eine genaue Bestimmung der eingesetzten Technologien nicht möglich, insbesondere in Fällen in denen proprietäre Algorithmen eingesetzt werden und eine Einsicht in den spezifischen Code nicht möglich ist. Da es sich bei dem Forschungsgegenstand dieser Arbeit um die gesellschaftlichen und ethischen Implikationen von KI handelt, sind die Auswirkungen dieser Technologien bedeutender als die eingesetzte Technologie selbst. Obwohl Weizenbaums ELIZA, im Kern eine Liste von einfachen Wenn-Dann-Regeln in Kombination mit variablen Textbausteinen, im Vergleich zu modernen KI-Technologien technisch äußerst simpel erscheint, inspirierte sie Weizenbaum und andere zeitgenössische Theoretiker und Praktiker zum Nachdenken über KI. Nichtsdestotrotz spielt die eingesetzte Technologie eine signifikante Rolle, wie in den späteren Kapiteln gezeigt werden soll. Insbesondere die Entwicklung und Verbreitung non-linearer Methoden wie Neuronale Netzwerke stellt einen kategorischen Unterschied zu vorher eingesetzten Methoden dar, in Hinsicht auf Transparenz der Methode und Nachvollziehbarkeit der Ergebnisse. Diese technische Entwicklung soll gezeigt und die Implikationen dargestellt werden, aber nicht zum primären Forschungsgegenstand werden.

3.2 Historische Entwicklung von KI-Technologien

In der Entwicklungsgeschichte der KI-Technologien hat es in den 65 Jahren seit der Dartmouth Conference im Jahr 1956 mehrere Höhe- und Tiefpunkte gegeben, mit einer Renaissance seit den frühen 2010er-Jahren. Um zu verstehen, warum die Frage nach den ethischen Implikationen von Künstlicher Intelligenz gerade heute wieder besonders relevant ist, ist eine kurze Betrachtung dieser Entwicklungsgeschichte sinnvoll.

In den 2010er-Jahren führten mehrere, unabhängige Trends in Computertechnologie zu einem signifikanten Sprung von KI-Technologien. “Big Data” beschreibt den Prozess des Sammelns, Speicherns und Analysierens von enormen Mengen von Daten. Obwohl dieser Begriff bereits seit den 1990ern benutzt wird, wurde die technologische Reife erst ab den 2000er-Jahren erreicht. Zum einen, weil die Entwicklung digitaler Speichermöglichkeiten die Speicherung von großen Mengen Daten sowohl technisch als auch finanziell möglich

gemacht hat, zum anderen weil die Menge an digital produzierten Daten durch soziale Netzwerke, Internet-of-Things-Sensoren und andere Quellen, signifikant gestiegen ist (LeCun et al. 2015). Parallel zu der Entwicklung von “Big Data”, wurden neue “Deep Learning”-Methoden im Bereich des Maschinenlernens entwickelt. “Deep Learning”-Methoden sind Algorithmen die auf künstlichen neuronalen Netzen (ANN, artifical neural network) basieren. Künstliche neuronale Netze sind inspiriert von der Funktionsweise biologischer Gehirne und nutzen mehrere Schichten künstlicher Neuronen. Jedes künstliche Neuron hat ein sogenanntes Gewicht, das die Stärke des eingehenden Signals repräsentiert, und eine Grenze, die definiert ab welcher Stärke das Neuron aktiviert wird. Ein ANN besteht aus mindestens drei Schichten: der “input layer” fungiert als erste Schicht und nimmt Daten an, die an einen oder mehrere “hidden layers” übergeben werden, wo die eigentliche Komputation stattfindet. Am Ende steht ein “output layer”, der das Ergebnis der Berechnung ausgibt. Die Funktionen dieser Schichten kann am Beispiel von Handschrift-Erkennung dargestellt werden. Der erste Schritt besteht darin, eine handschriftliche Zahl zu einem Raster-Bild zu transformieren, in dem jeder Pixel ein künstliches Neuron darstellt und die Helligkeit des Pixels definiert das Gewicht des Pixels – dies ist der input layer. In den hidden layers wird sich anschließend schrittweise die Zahl erkannt: beispielsweise wird erst im oberen Bereich ein Kreis erkannt (dies könnte auf eine 8, aber auch eine 9 hindeuten), danach wird im unteren Bereich des Bildes ein weiterer Kreis erkannt, was die Wahrscheinlichkeit für eine “8” gegenüber einer “9” massiv steigert. Das Ergebnis, dass eine “8” erkannt wurde, wird schlussendlich im output layer als Ergebnis ausgegeben.

Diese neuronalen Netzwerke stellen in der Regel “Black Boxes” dar, da da die “hidden layers” nicht beobachtbar sind und deswegen nicht transparent ist, warum ein Netzwerk zu einem bestimmten Resultate gekommen ist. Die Entwicklung dieser “Black Box”-Modelle die auf Methoden des maschinellen Lernens bestehen haben große Implikationen für die Erklärbarkeit der produzierten Resultate. Dies liegt daran, dass die Berechnungen des Algorithmus zwischen Eingabe der Daten und dem Resultate völlig unklar bleiben und keine Erklärung liefern, die für Menschen interpretierbar sind. Dies liegt unter anderem daran, dass diese Modelle teilweise hunderte Millionen Parameter in ihre Berechnungen einfließen lassen, die unmöglich von Menschen nachzuvollziehen sind (vgl. Buhrmester et al. 2019). Allgemein hat diese Komplexität den Vorteil, dass sie Algorithmen mit höherer

Präzision produziert; je besser die Vorhersagekraft des Modells, desto schlechter wird seine Erklärbarkeit (vgl. Gunning 2017). Dies bedeutet nicht unbedingt, dass diese komplexen “Black Box”-Modelle besser funktionieren, als andere Modelle. So konnte gezeigt werden, dass ein einfacheres Decision-Tree-Modell vergleichbar gut funktioniert wie das komplexe Black-Box-Modell von COMPAS, einer Software die die Rückfallraten von Straftätern vorhersagt (Angelino et al. 2017). Probleme mit der Erklärbarkeit können in vielen Formen auftreten. So kann es sein, dass ein Modell nur aus Zufall richtige Ergebnisse produziert. Ein Beispiel dafür ist ein Algorithmus, der Huskies und Wölfe unterscheiden soll. Da auf den Bildern mit denen das Modell die Klassifizierung gelernt hat, Wölfe fast immer vor einem Hintergrund mit Schnee zu sehen war, lernte der Algorithmus in Wirklichkeit die Präsenz von Schnee zu klassifizieren (Ribeiro et al. 2016). Ein weiteres Problem bei der Entwicklung von KI-Anwendungen spielen Daten eine maßgebliche Rolle, da Klassifizierungs-Algorithmen von ihnen lernen, die richtigen Vorhersagen zu machen. In diesen Datensets kann bereits ein bias existieren, also eine Verzerrung unterschiedlicher Form existieren. Da sich die Verarbeitung dieser Daten nicht nachvollziehen lässt, kann dieser bias unerkannt bleiben.

3.3 Zusammenfassung

Es lassen sich mehrere Trends in der KI-Entwicklung und -Forschung feststellen, die im Kontext dieser Arbeit relevant sind. Die Entwicklung von linearen zu non-linearen Methoden wie Deep Learning und neuronalen Netzwerken hat zu Black-Box-Modellen geführt, die sich einer Prüfung und Interpretierbarkeit entziehen. Diese Algorithmen aus dem Bereich des Deep Learnings können als das betrachtet werden, was Weizenbaum als “nicht theorie-basierte” Programme betrachtet; ihre internen Zuständen sind für externe Betrachtung ausgeschlossen (vgl. Weizenbaum 1966, 247). Dieser technische Hintergrund macht die im nächsten Kapitel beschriebenen ADM-Methoden verständlicher.

4 Algorithmische Entscheidungsfindung (ADM)

Algorithmische Entscheidungsfindung (algorithmic decision-making, ADM) ist im Kontext der bisher ausgearbeiteten Weizenbaumschen Medientheorie von besonderer Bedeutung da sie eine der größten und relevantesten Überschneidungen zwischen alltäglichem Leben und neuartigen KI-Technologien darstellt. Obwohl es weder die Begrifflichkeit “algorithmische Entscheidungsfindung” noch die nötigen Technologien gab, als Weizenbaum 1976 “Computer Power and Human Reason” veröffentlichte, befasste er sich dort bereits mit der Frage, welche Funktionen, beziehungsweise gesellschaftliche Rollen, von Computern übernommen werden dürfen. Es ist genau im Bereich der algorithmischen Entscheidungsfindung, dass vormals gesellschaftlich relevante Entscheidungen – wie die Vergabe von Therapieplätzen oder die Verurteilung von Straftätern – von Maschinen ersetzt werden. In diesem Kapitel soll erklärt werden, was algorithmische Entscheidungsfindung ist und welche Probleme beim Einsatz dieser Systeme auftreten.

4.1 Was ist algorithmische Entscheidungsfindung?

Algorithmische Entscheidungssysteme werden bereits in vielen Bereichen des privaten und öffentlichen Lebens eingesetzt. Algorithmisch getroffene Entscheidungen können im gesellschaftlichen Zusammenleben sowohl auf persönlicher als auch auf öffentlicher Ebene einen signifikanten Einfluss haben (Lischka und Klingel 2017). So setzen Unternehmen diese Systeme ein um die potentielle Produktivität von Angestellten zu berechnen um basierend auf diesen Resultaten Bewerber einzustellen oder nicht (Chalfin et al. 2016). Auch bei der Vergabe von Krediten werden ADM-Systeme genutzt, um die Kreditwürdigkeit von Bewerbern zu berechnen (Khandani et al. 2010). In der Kriminologie werden unter der Bezeichnung “predictive policing“ Systeme entwickelt, die Muster im Vorgehen einzelner Krimineller und Gruppen von Kriminellen erkennen um vorherzusagen, wo diese als nächstes aktiv werden (Wang et al. 2013). Nicht nur in der Polizeiarbeit, auch im Justizwesen werden bereits ADM-Systeme eingesetzt; ein Beispiel dafür ist der Einsatz eines Systems im US-Bundesstaat Pennsylvania. In der Hoffnung, die Kriminalität zu senken als auch Gefängnisse zu entlasten, produziert dieses System

basierend auf einer Risikoeinschätzung eine Empfehlung für Richter, wie lange die Haftstrafe für einen verurteilten Straftäter sein sollte (Barry-Jester et al. 2015). Im Bildungswesen werden im Rahmen von “learning analytics” Systeme eingesetzt, die Universitäts-Bewerber automatisiert klassifizieren und darüber entscheiden, welche Studenten an einer Universität zugelassen werden (Jones & McCoy 2019).

Allen diesen Systeme haben das, möglichst exakte Vorhersagen zu treffen und damit menschliche Entscheidungsfindung zu unterstützen (Kleinberg et al. 2015). Auf den ersten Blick bietet der Einsatz von KI-Systemen zahlreiche Vorteile: Sie sind schneller als Menschen und können eine größere Menge Daten einbeziehen als es für Menschen möglich wäre. Gleichzeitig scheinen algorithmische Entscheidungen potentiell objektiver als die eines Menschen; als deterministische Maschinen produzieren sie bei gleichen Daten das gleiche Resultat und lassen sich nicht von den zahlreichen systematischen Fehlern beeinflussen, die menschliches Urteilsvermögen inakkurat machen (Tversky und Kahneman 1974).

4.2 Probleme algorithmischer Entscheidungsfindung

Obwohl ADM-Systeme vermeintlich objektive Vorhersagen produzieren, sind diese Vorhersagen das Produkt eines komplexen Prozesses bei dem bei jedem Schritt ethische Entscheidungen getroffen werden müssen, die gesellschaftliche Auswirkungen haben können. Lischka und Klingel (2017) haben eine Liste von Fallbeispielen von bereits existierenden ADM-Systemen zusammengestellt, aus denen sich konkrete Probleme und Herausforderungen für den Einsatz dieser Systeme ableiten lassen.

1.) Falsifizierbarkeit des Algorithmus: Grundsätzlich können sich ADM-Systeme durch Feedback-Schleifen und neue Daten weiterentwickeln um die Genauigkeit ihrer Vorhersagen zu verbessern. Die Möglichkeiten dieser Weiterentwicklung sind jedoch eingeschränkt, da nur tatsächlich eingetroffene Zustände an den Algorithmus zurückgespielt werden können. So kann beispielsweise bei einer Kreditvergabe nachvollzogen werden, ob jemand, der vom Algorithmus als kreditwürdig kategorisiert wurde, den Kredit tatsächlich zurückgezahlt hat oder nicht. Gleichzeitig können als nicht

kreditwürdig eingestufte Personen nicht beweisen, dass sie den Kredit doch zurückgezahlt hätten. So entsteht “asymmetrisches Feedback”, das die Weiterentwicklung des Algorithmus einschränken oder sogar verfälschen kann (vgl. Zweig & Kraft 2018, 221f).

2.) Sachgerechte Anwendung des Algorithmus: Es muss sichergestellt werden, dass das ADM-System nur für das Ziel eingesetzt wird, für das es entwickelt wurde. Eine Missachtung dieser Regel kann beispielsweise dazu führen, dass die Resultate eines Algorithmus zur Kriminalitätsprognose bereits zu einer Vorverurteilung führen.

3.) Richtigkeit der Wirkungslogik: Die hohe technische Effizienz von ADM-Prozessen kann dazu führen, dass ihr Einsatz gegenüber anderen Lösungsmöglichkeiten privilegiert wird wodurch andere, nicht-technische Möglichkeiten nur noch eingeschränkt in Betracht gezogen werden. Zusätzlich kann dies zur Folge haben, dass andere Ansätze zur Lösung des Problems in ihrem Umfang reduziert oder sogar abgeschafft werden. So ging in Chicago die Einführung eines Algorithmus zur Bestimmung von Wohnhäusern mit hohem Risiko für Bleivergiftung mit einer Reduzierung der zuständigen Kontrolleure einher (vgl. Lischka und Klingel 2017, 14).

4.) Operationalisierbarkeit des Problems oder Phänomens: Operationalisierung beschreibt den Prozess, bestimmte Parameter eines Problems oder Phänomens, zu dem es keinen direkten Zugang gibt, zu definieren und verwertbar zu machen. Im Kontext von ADM-Systemen bedeutet dies zum Beispiel, den zu prognostizierenden Wert zu definieren sowie die Parameter, die in die Vorhersage dieses Wertes einfließen sollen. Die Herausforderung besteht darin, dass diese Operationalisierungen in einem öffentlichen Diskurs entwickelt werden um sicherzustellen, dass diese fachlich untermauert sind und gesellschaftliche Interessen widerspiegeln (vgl. Lischka und Klingel 2017, 17).

5.) Evaluation der Auswirkungen: ADM-Systeme ermöglichen die Nutzung und Verarbeitung von enormen Datenmengen in einer Art und Weise, die vorher mit manuellen Methoden nicht möglich gewesen wäre. Dadurch wird es nötig, dass diese Einsatzszenarien einer permanente Evaluation unterzogen werden, die den individuellen

und gesellschaftlichen Anforderungen an den Algorithmus gerecht werden (vgl. Lischka & Klingel, 2017, 18f).

6.) Vielfalt von Algorithmen: Gleiche oder ähnliche ADM-Algorithmen lassen sich in der Regel auf mehrere unterschiedliche Fälle und Problemfelder anwenden. Dies kann potenziell dazu führen, dass ein einzelner leistungsstarker Algorithmus benutzt wird und dessen Funktionsweise zum de facto Standard wird. Wenn dieser Algorithmus bestimmte systematische Fehler beinhaltet, würden diese in vielen Bereichen gleichzeitig erscheinen. Dieses Problem kann durch den Einsatz von einer Vielfalt von Algorithmen abgeschwächt werden – mehrere Algorithmen in ähnlichen Bereichen konkurrieren zu lassen würde garantieren, dass kein einzelner Algorithmus Überhand nimmt (vgl. Lischka und Klingel 2017, 23).

7.) Überprüfbarkeit des Algorithmus: Ein Problem von Algorithmen, insbesondere proprietärer, ist, dass ihre Funktionsweise, die Auswahl ihrer Parameter und ihr tatsächlicher Quellcode nicht überprüfbar sind. Lischka & Klingel (2017) beschreiben die Auswirkungen dieses Problems anhand des ADM-Systems “Admission Post Bac” (APB), welches seit 2009 für die Studienplatzvergabe französischer Abiturienten eingesetzt wird. Dieses System verteilt begrenzte Studienplätze nicht nur auf Basis der expliziten Präferenzen der Abiturienten, sondern lässt auch die Nähe des Wohnorts zur Wahl-Universität miteinfließen. Da viele begehrte Elite-Hochschulen in Frankreich in teuren Städten wie Paris liegen, führt dieser Faktor zu einer systematischen Benachteiligung von Bewerbern außerhalb von Paris und anderer teurer Städte. Dadurch führt der Faktor des Wohnorts zu einer impliziten soziokönomischen Selektion durch den Algorithmus. Erst 2016, sieben Jahre nach Einführung des Systems, wurde diese Selektion bekannt, nachdem eine Klage dazu führte dass die französische Regierung den Code offenlegen musste. Durch das Gerichtsurteil konnten erstmals Dritte den Code überprüfen um diese implizite Selektion darzustellen und öffentlich zu diskutieren. Dieses Fallbeispiel zeigt deutlich, warum ADM-Systeme überprüfbar sein sollten – nur durch die Überprüfung durch unabhängige Dritte kann garantiert werden, dass Algorithmen keine systematischen Verzerrungen beinhalten (vgl. Lischka und Klingel 2017, 25f).

8.) Soziale Wechselwirkungen: Der Einsatz von ADM-Systemen stellt einen Feedback-Loop mit sozialen Systemen dar. Dies wird im insbesondere im Bereich des Predictive Policing deutlich, in dem Systeme zum Einsatz kommen die die Wahrscheinlichkeit von kriminellen Handlungen in einem bestimmten Gebiete, beispielsweise Wohnungseinbrüche, vorhersagen. Obwohl es Indikatoren gibt dass diese Systeme tatsächlich besser als Menschen Kriminalität vorhersagen und verhindern können, sind die sozialen Wechselwirkungen unklar. Darunter fällt zum Beispiel die mögliche Wechselwirkung dass Täter ihre Handlungen an das genutzte System anpassen oder ihre Einsatzorte wechseln und das ADM-System die Kriminalität letztendlich nur verdrängt. Gleichzeitig kann der Einsatz des Systems zu einem verzerrten Fokus auf Verbrechen aus dem Hellfeld führen, da der Algorithmus nicht mit nicht-existenten Daten aus dem Dunkelfeld gefüttert werden kann. Diesem Problem muss mit einer ständigen Evaluierung der Auswirkungen begegnet werden, um negative Wechselwirkungen zu vermeiden (vgl. Lischka und Klingel, 2017, 28f).

9.) Zweckentfremdung des Systems: Resultate eines ADM-Systems können potentiell auch zu Beurteilungen genutzt werden, für die sie überhaupt nicht entwickelt wurden. Lischka & Klingel zeigen dies anhand der Berechnung von Kreditwürdigkeit in den USA. Das Scoring der Kreditwürdigkeit sollte eigentlich nur für die Entscheidung genutzt werden, einen Kredit an eine Person zu vergeben oder nicht. Mittlerweile hat sich jedoch die Praxis eingestellt, diesen Wert auch für viele andere Entscheidungen einzusetzen, beispielsweise bei der Berechnung von Versicherungshöhen oder sogar, wie lange Kunden in einer Telefon-Warteschlange warten müssen. Dabei bleibt zweifelhaft, wie sinnvoll die Nutzung der Kreditwürdigkeitsbewertung für andere Zwecke wirklich ist. An diesem Beispiel wird deutlich, dass durch organisatorische Bequemlichkeit und vermeintlichem Effizienzgewinn durch die mehrfache Nutzung des gleichen Resultats, eine negative Bewertung in einem Bereich zu negativen Folgen in anderen Lebensbereichen führen kann (vgl. Lischka und Klingel 2017, 31f).

Diese Fallbeispiele lassen sich in zwei Gruppen von Problemen unterteilen. In der ersten Gruppe befinden sich Probleme, die aus den Eigenschaften des Algorithmus beziehungsweise der technischen Prozedur entstehen. Hierzu gehören die Probleme der

Falsifizierbarkeit und der Überprüfbarkeit des Algorithmus. Die zweite Gruppe von Problemen lässt sich dahingehend zusammenfassen, dass sie ein Resultat davon sind, in welcher organisatorischen Form der Algorithmus eingesetzt wird. Zu dieser Gruppe gehören die sachgerechte Anwendung des Algorithmus, die Richtigkeit der Wirkungslogik, die Evaluation der Auswirkungen, der Einsatz vielfältiger Systeme, die Berücksichtigung sozialer Wechselwirkungen und der Schutz vor Zweckentfremdung. Zwischen diesen beiden Gruppen befindet sich die Problematik der Operationalisierung. Diese Problematik ergibt sich sowohl aus dem Organisationsproblem der Entwicklung der Parameter als auch aus den technischen Limitierungen, inwiefern diese Parameter innerhalb eines Algorithmus operationalisierbar sind.

In mehreren Fallbeispielen wird der menschliche Einfluss auf den vermeintlich objektiven Algorithmus besonders deutlich, insbesondere bei Fragen der Operationalisierung von komplexen sozialen Phänomenen und ihrer Übersetzung in eine Form von Vorhersage-Problem, für das ADM-Systeme geeignet sind. Bei diesem Entwicklungsprozess werden mehrere Schritten durchlaufen, bei denen menschliche Entscheidungen die Qualität und Ausrichtung des finalen Algorithmus beeinflussen. Darunter fallen nach Zweig und Krafft (2018):

1.) Datenauswahl: Die Entwickler eines ADM-Systems müssen Entscheidungen treffen, mit welchen Daten der Algorithmus trainiert werden soll. Diese Daten müssen idealerweise vollständig genug und in einem geeigneten Kontext erhoben worden sein, um eine sinnvolle Basis für den Algorithmus darzustellen. Verzerrungen in der Datenauswahl sind besonders schwerwiegend, da sie die späteren Vorhersagen des Algorithmus fundamental beeinflussen.

2.) Quantifizierung, beziehungsweise Operationalisierung: Werden ADM-Systeme zur Unterstützung komplexer sozialer Phänomene und Herausforderungen eingesetzt, muss zwangsweise eine Selektion vorgenommen werden, welche Aspekte dieses Phänomens als Parameter in einem Algorithmus benutzt werden sollen. Diese Auswahl ist zutiefst sozial, kulturell und ethisch geprägt. Dies zeigt sich beispielsweise an der Studienplatzvergabe in Frankreich, bei der eine kulturelle Annahme – dass Abiturienten

mehr Recht auf einen Studienplatz an einer Universität haben, wenn sie bereits in der Nähe dieser Universität leben – dazu geführt hat, dass durch den Algorithmus eine implizite soziale Selektion vorgenommen wurde.

3.) Methode des maschinellen Lernens: Wie im Kapitel “Künstliche Intelligenz” beschrieben, unterscheiden sich Methoden des maschinellen Lernens nicht nur in ihren technischen Eigenschaften und Möglichkeiten. Die eingesetzte Methode hat vielmehr einen signifikanten Einfluss auf die Transparenz und Erklärbarkeit der Methode und ihrer Resultate.

4.) Kriterien für Qualität und Fairness der Methode: Qualität und Fairness bedeuten im Kontext von ADM-Systemen, wie gut das System im alltäglichen Einsatz funktioniert und ob diese Entscheidungen fair getroffen werden, also nicht diskriminiert. Eine einfache Beurteilung des Qualitätsmaß besteht darin, die jeweils richtigen (“true positives” und “true negatives”) und falschen (“false positives” und “false negatives”) Klassifizierungen eines Systems zu summieren. Doch auch diese vermeintlich simple Summierung, die beispielsweise zeigen könnte dass ein System tatsächlich mehr “true positives” als “false positives” produziert, verschleiert dahinter liegende soziale Entscheidungen. Die reine numerische Berechnung lässt die unterschiedlichen gesellschaftlichen Bewertungen außer Acht, konkret ob zum Beispiel die korrekte Vorhersage dass ein Straftäter rückfällig wird (true positive) wichtiger ist als einem Straftäter zu Unrecht eine hohe Rückfallwahrscheinlichkeit vorherzusagen (false negative). Auch diese Bewertungen sind sind kulturell und sozial geprägt und lassen sich nicht nach objektiven Maßstäben beurteilen (vgl. Zweig und Krafft 2018, 211f).

5.) Kontext des Einsatzes eines ADM-Systems: Am Ende des Entwicklungsprozess des ADM-Systems stehen Entscheidungen, in welchem Kontext und auf welche Art das System eingesetzt wird. In diesem Schritt befinden sich viele der Herausforderungen, die sich in den Fallbeispielen von Lischka und Klingel (2017) als Organisationsprobleme einordnen lassen. Außerdem müssen hier wichtige Entscheidungen getroffen werden, wie autonom das System ist – ob es Entscheidungen völlig selbst treffen darf, ob es nur unterstützend eingesetzt wird und in welchem Maße menschliche Entscheider eingreifen

können. Auch eine Rolle spielt, wie gut Endnutzer das System verstehen und in ihre Entscheidungen einbetten können.

Weizenbaums Kritik bezieht sich insbesonderen auf den Schritt der Operationalisierung und Quantifizierung sozialer Fragen zum Einsatz in ADM-Systemen. Diesen Prozess sieht er nicht als technische Herausforderung, vielmehr besteht für ihn eine fundamentale Unmöglichkeit, bestimmte Prozesse operationalisierbar zu machen. Nach Weizenbaum gibt es eine Form von Wissen, die sich eine Person während ihrer Sozialisation und “mit ihrer Muttermilch”, also von Generation zu Generation, aneignet. Dieses Wissen beinhaltet ein intuitives Verständnis bestimmter kultureller Normen und Regeln die sich nicht mit Büchern erlernen lassen und die in keiner anderen Form als “dem Leben selbst” ausgedrückt werden können. Dieses kulturelle Wissen beschreibt Weizenbaum am Beispiel von kulturell unterschiedlicher Rechtsprechung. Da die amerikanische und japanische Kultur so unterschiedlich ist, so Weizenbaum, könnte ein amerikanischer Richter niemals an einem japanischen Familiengericht sprechen: “An American judge, therefore, no matter what his intelligence and fairmindedness, could not sit in a Japanese family court” (vgl. Weizenbaum 1976, 226). Weizenbaum beschreibt damit auch eine Form von Wissen, die sich grundsätzlich nicht formalisieren lässt und damit auch niemals für ein Computerprogramm operationalisiert werden kann (vgl. Weizenbaum 1976, 225). Da keiner künstlichen Intelligenz jemals diese Form von intuitivem Wissen beigebracht werden kann, betrachtet Weizenbaum sie als grundsätzlich ungeeignet für Aufgaben, für die dieses Wissen nötig ist. Gleichzeitig schließt Weizenbaum nicht aus, dass Computersysteme auch in diesen Fragen zu korrekten Antworten kommen können; trotzdem sollten diese Antworten abgelehnt werden, da ihre Herleitung grundlegend falsch ist: “They may even be able to arrive at ‘correct’ decisions in some cases – but always and necessarily on bases that no human being should ever be willing to accept.” (Weizenbaum 1976, 227).

Ein konkretes Beispiel für ein Operationalisierungs-Problem stellt die Operationalisierung von Fairness dar; dies soll am Beispiel der Software COMPAS gezeigt werden. COMPAS wird in den USA eingesetzt um das Rückfallrisiko von Straftätern zu berechnen und diese in verschiedene Risikogruppen einzuteilen. Die Software wurde dafür kritisiert, dass

schwarze Straftäter, die nach ihrer Entlassung nicht rückfällig wurden, trotzdem vorher systematisch mit einem höheren Risiko eingestuft wurden (Angwin et al. 2016). In diesem Kontext stehen sich zwei Definitionen von Fairness gegenüber: individuelle Fairness und Gruppen-Fairness. Beide Definitionen sind mathematisch unvereinbar wenn sich das Rückfallrisiko zwischen beiden Gruppen tatsächlich unterscheidet, wie im folgenden gezeigt werden soll (vgl. Tolan 2019). Fairness wird im folgenden definiert als die gleiche Behandlung unterschiedlicher Personen unabhängig davon, ob sie einer bestimmten Gruppe angehören. Personen können anhand von “legitimen” Eigenschaften mit hoher Vorhersagekraft klassifiziert werden, also beispielsweise Anzahl von früheren Straftaten, aber nicht anhand von “geschützten” Eigenschaften wie ihrer Ethnie. Gruppen-Fairness bedeutet im Kontext von Rückfallrisiko, dass der Anteil von Hochrisiko-Klassifizierungen in allen Gruppen den gleichen Anteil haben sollte. Dies bedeutet, dass Mitglieder unterschiedlicher Gruppen bei gleichen legitimen Eigenschaften die gleiche Klassifizierung erhalten sollten. Wenn sich das tatsächliche Risiko zwischen den Gruppen aber unterscheidet, führt dies zu Problemen. So müssten, um die Risikoverteilung gleich zu halten, genug Mitglieder der Gruppe mit dem niedrigeren Risiko höher eingestuft werden, auch wenn diese sonst eigentlich niedriger eingestuft worden wären. Alternativ müssten Personen mit hohem Risiko mit einem niedrigen Risiko eingestuft werden, was wiederum dazu führen kann dass diese Gruppe hohe Rückfallraten produziert (Dwork et al. 2012). Bei individueller Fairness wird der Fokus auf individuelle Personen und deren Eigenschaften gelegt und das Rückfallrisiko nur auf Basis dieser individuellen Parameter berechnet. Diese Definition wirft die Frage auf, welche Parameter für die Klassifizierung genutzt werden sollen und genutzt werden dürfen. Denn auch wenn geschützte Attribute wie Geschlecht oder Ethnie nicht genutzt werden, können sie in der Berechnung dennoch eine Rolle spielen wenn diese Attribute stark mit anderen legitimen Parametern (beispielsweise sozioökonomischen Variablen) korrelieren (Barocas und Selbst 2016). Bei tatsächlichen Unterschieden zwischen zwei Gruppen, zum Beispiel einem insgesamt niedrigerem Rückfallrisiko von Frauen gegenüber Männern, würden ohne Geschlecht als Kriterium Frauen systematisch mit einem zu hohen Risiko klassifiziert werden. Um dies auszugleichen, müssten Frauen und Männer unterschiedliche Werte haben, ab denen sie ein hohes Risiko erhalten. Damit würden aber Personen an unterschiedlichen Standards gemessen werden, für die sie keine Verantwortung tragen, zum Beispiel ihr Geschlecht (Chouldechova (2017). Es wird deutlich, dass ein Computersystem, egal wie schnell oder

vermeintlich intelligent, könnte demnach niemals ein Entscheidung produzieren, der alle Definitionen von Fairness in einem bestimmten Kontext genügen. Dies liegt an der Limitierung des Mediums Computers, dessen Operationen in jedem Fall auf einer Quantifizierung von “Fairness” bestehen müssen. An diesem Beispiel wird der von Weizenbaum beschriebene Konflikt zwischen menschlicher Intuition und künstlicher Intelligenz besonders deutlich: das intuitive Verständnis, das Menschen von Fairness besitzen, lässt sich logisch unmöglich mathematisch formalisieren und so von Computern berechnen lassen. An diesem Beispiel wird deutlich, welchen Gedanken Weizenbaum bei der Unterscheidung von Entscheiden (“deciding”) und Wählen (“choosing”) hatte. Wenn es um Fragen der Fairness geht, kann der Computer eine Entscheidung treffen, also auf Basis programmierter Parameter und Daten eine Klassifizierung vornehmen. Was der Computer aber wegen der mathematischen Unvereinbarkeit nicht vornehmen sollte, ist eine Wahl zwischen verschiedenen Fairness-Definitionen, da diese Wahl in hohem Maße kulturell und gesellschaftlich geprägt ist. Hinter diesen technischen Problemen stehen ethische und gesellschaftliche Abwägungen und Fragen nach öffentlichen Gütern und Werten. Damit zeigt das Problem der Operationalisierung von Fairness, dass es keine funktionierende Formalisierung von Fairness außerhalb ethischer Abwägungen innerhalb einer spezifischen Situation geben kann.

In diesem Kapitel wurde dargestellt, dass hinter den vermeintlich objektiven Resultaten eines algorithmischen Entscheidungssystems in Wirklichkeit eine Vielzahl menschlicher Entscheidungen steckt. Bereits in der Entwicklung eines Algorithmus werden implizit oder explizit Entscheidungen getroffen, die sozial und kulturell beeinflusst sind. Dies beginnt mit der Auswahl und Verfügbarkeit der Trainingsdaten, besteht in der Art und Weise wie ein soziales Phänomen quantifiziert wird und endet mit dem sozialen Kontext, in dem das System eingesetzt wird und wie dessen Resultate in einem größeren Zusammenhang genutzt werden. Außerdem wird deutlich, dass bestimmte Operationalisierungen, zum Beispiel von Fairness, mathematisch unvereinbar sind und die Entscheidung darüber, welche Operationalisierung genutzt werden soll, auf Basis ethischer Abwägungen passieren muss. Einige dieser Aspekte wurden bereits von Weizenbaum in “Computer Power and Human Reason” angesprochen. Der Eigenschaften des Computers als Medium schränkt bereits im Voraus die Auswahlmöglichkeiten des Algorithmus ein. Ein Algorithmus basiert zwangsweise auf einem mathematischen Modell und kann nicht von

sich aus unterschiedliche Bewertungen einfließen lassen wie beispielsweise ein menschlicher Richter. Das Beispiel der Fairness zeigt außerdem, dass nicht alles menschliches Wissen, in diesem Fall über Fairness und die richterliche Einschätzung von Straftätern, in einem Computersystem formalisiert werden kann.

Diese Probleme von ADM-Systeme sind KI-Forschern und -Praktikern bekannt. Im Folgenden sollen Lösungsansätze erläutert werden, die sich mit diesen Problemen befassen. Um mit den Problemen umzugehen, die durch den Einsatz von ADM-Systemen entstehen, gibt es mehrere unterschiedliche Lösungsansätze. Diese setzen entweder bei der Funktionsweise des Algorithmus selbst an oder regulieren die Einsatzmöglichkeiten der Systeme.

Um die Akzeptanz von algorithmischen Entscheidungen zu verstärken, werden Algorithmen entwickelt die sowohl transparent als auch erklärbar sind und damit Transparenz schaffen (Sample 2017). Daneben gibt es rechtliche Ansätze, zum Beispiel mit der Datenschutzgrundverordnung (DSGVO) der Europäischen Union, die gewisse Rechte gegenüber algorithmisch getroffenen Entscheidungen garantieren. Ein Rechtsanspruch auf Erklärbarkeit setzt jedoch technische Methoden voraus, die Algorithmen transparent und ihre Entscheidungen erklärbar machen. Obwohl die Entwicklung von Methoden, die die Entscheidungen von Algorithmen transparent und erklärbar machen, ein aktives Forschungsfeld ist, sind diese Methoden weder universal verfügbar noch werden sie flächendeckend eingesetzt (Samek und Müller 2019). Mit der Datenschutzgrundverordnung der EU (DSGVO / GDPR) wurde mit dem “Recht auf Erklärung” eine rechtliche Basis für Individuen geschaffen, die das Subjekt algorithmischer Klassifizierung sind und eine Erklärung für die Klassifizierung erhalten wollen. Bei dieser Frage stehen häufig die Interessen der profilierten Personen den Interessen der KI-Entwickler gegenüber, beispielsweise wenn es um das berechtigte Interesse geht, Betriebsgeheimnisse zu schützen und die genutzten Trainingsdaten vor öffentlichem Zugriff zu schützen. Dem gegenüber steht das Interesse der betroffenen Personen, zu verstehen warum sie in einer bestimmten Form klassifiziert wurden, auf Basis welcher

Parameter und unter welchen Umständen sie anders klassifiziert worden wären. So könnte zum Beispiel eine Privatperson, deren Bewerbung um einen Kredit abgelehnt wurde, besser verstehen, nach welchen Änderungen sie vom Algorithmus kreditwürdiger erscheint. Eine mögliche Lösung für die Umsetzung dieses Rechtsanspruch sind “unconditional counterfactual explanations”, die die kleinstmögliche Input-Veränderungen beschreiben, mit der das gewollte Resultate annähernd produziert werden kann (Wachter et al. 2017).

Doch an dem Ansatz, ADM-Systeme rechtlich zu regulieren, gibt es Kritik. Die Idee, diese Systeme rechtlich zu regulieren, geht von der impliziten Annahme aus, dass eine Erklärung des produzierten Resultats in jedem Fall möglich ist. Dies ist allerdings weder aktuell, noch für die Zukunft, gegeben. Durch den Einsatz von Black-Box-Modellen würden KI-Systeme sich grundsätzlich klassischen Formen der Regulierung entziehen, da ihre inneren Zustände nicht beobachtbar sind (Bathaee 2017). Dazu kommt, dass Rechtsmittel erst eingesetzt werden können, nachdem Unrecht passiert ist, beispielsweise eine falsche Klassifizierung. Zusätzlich ist es fragwürdig, mit welchem Ansatz Betroffene gegen Entscheidungen klagen könnten. So Schreiben Lischka & Klingel abschließend: “Rechtsmittel scheinen bei diesen Fällen derzeit ein schwaches Korrektiv zu ein. Denn die Intransparenz der Verfahren und Entscheidungslogiken erschwert es Bewerteten, einen Ansatzpunkt für eine Beschwerde oder gar Klage zu entdecken” (Lischka & Klingel 2017, 23).

4.3 ADM: Fazit

Es wird deutlich, dass ADM-Systeme heute bereits eine enorme Rolle im gesellschaftlichen Zusammenleben spielen und damit einen großen Einfluss auf einzelne Menschen als auch die Gemeinschaft haben. Diese Systeme werden wegen ihres Versprechens großer Effektivitätsgewinne eingesetzt. So erlauben ADM-Systeme die Verarbeitung großer Datenmengen aus denen Rückschlüsse und Prognosen entwickelt werden können, die mit manuellen oder analogen Mitteln nicht möglich sind. Doch ihre überragende technische Effizienz produziert eine Reihe von Herausforderungen und Problemen.Neben ihrer technischen Effizienz versprechen diese Systeme eine

maschinelle Objektivität, die Entscheidungen frei von den Fehlern produziert, die natürlicher Teil menschlichen Entscheidungsprozesse sind. Nichtsdestotrotz wird evident, dass auch diese vermeintlich objektiven Entscheidungen beeinflusst sind von den gleichen Faktoren, die auch bei Menschen zu Fehlern in ihrer Entscheidungsfindung führen. Da diese menschlichen Entscheidungen, beispielsweise die Auswahl der Daten und die Wahl einer Fairness-Klassifikation, inhärenter Teil des Entwicklungsprozesses eines ADM-Systems sind, ist die vermeintliche Objektivität von ADM-Systemen nur eine Illusion.

Die unterschiedlichen Ansätze ADM-Systeme zu regulieren, haben gemein, dass sie den Einfluss der Systeme nicht grundsätzlich reduzieren sondern lediglich berechenbarer machen. Transparenz, Erklärbarkeit und ein Recht auf Erklärung können ein Korrektiv gegen die schwerwiegendsten Auswirkungen dieser Systeme darstellen, aber es ist fraglich, ob sie selbst damit erfolgreich sind.

5 Zusammenfassung und Fazit

Das Ziel dieser Arbeit war es, Joseph Weizenbaums Diskussion von KI-Technologien aus dem Jahr 1976 medientheoretisch einzuordnen und in den Kontext moderner Entwicklungen zu setzen um die Frage zu beantworten, wie relevant Weizenbaums Kritik im 21. Jahrhundert ist. Viele von Weizenbaums Kritikpunkten bestehen so auch noch im Jahr 2020 und sind teilweise noch bedeutender geworden. Dies liegt an einem der zentralen Missverständnisse, wenn es um die Einordnung von Problemen mit ADM-Systemen geht: die mangelnde Unterscheidung die zwischen dem Computer als Medium und Algorithmen als Inhalt dieses Mediums gemacht wird. Die Ideen Weizenbaums und deren Einordnung in die Medientheorie von McLuhan hat deutlich gemacht, dass die Betrachtung des Mediums ertragreicher sind als die Betrachtung seines Inhalts, wenn es um die Diskussion der ethischen Implikationen algorithmischer Entscheidungsfindung geht. Erst der Fokus auf die Eigenschaften, Wirkungsmöglichkeiten und Einschränkungen des Mediums Computer macht eine tiefergehende Analyse möglich. Während McLuhan sich jedoch nicht ausführlich mit den Philosophien und Weltbildern hinter einem Medium beschäftigt und Entscheidungen für oder gegen ein Medium in Präferenzen zu bestimmten Medien-Formen sieht, ist für Weizenbaum der Computer die

Verkörperung eines reduktionistischen, quantitativen Weltbilds. Somit erweitert Weizenbaum die Medientheorie des Computers um die Betrachtung, dass ein bestimmtes Ideal von Wissenschaft und Rationalität eine maßgebliche Rolle für die Entwicklung neuer Computer-Technologien spielt. Im Gegensatz zu dieser Position wirkt die “ideologiefreie” Perspektive von McLuhan fast wie eine naive, technikfreundliche Position, die die tieferen Hintergründe des Computers verkennt. Zusätzlich fügte Weizenbaum die Unterscheidung zwischen prostethischen und autonomen Werkzeugen beziehungsweise Medien hinzu. Diese Unterscheidung verdeutlicht, dass Menschen Medien schaffen können, die nach einem ihnen fremden Modell der Welt arbeiten und ihnen dieses Modell aufzwingen können. Ein Punkt, in dem Weizenbaum und McLuhan übereinstimmen, ist dass technische Lösungen blind machen können für die größeren Auswirkungen des Mediums. Weizenbaum hat bereits in den 1970ern die Tendenz beschrieben, dass politische und ethische Probleme als technische Probleme umgedeutet werden um diese als vermeintlich lösbar darzustellen. In diesem Kontext erwähnt er die Reaktion von Universitäten auf Proteste gegen den Vietnamkrieg: der politische Widerstand gegen diesen wurde von den Universitäten als Kommunikationsproblem interpretiert und probiert, mit der Einrichtung von speziellen Hotlines zu lösen; eine technische Lösung für ein politisches Problem. Dieser Trend setzt sich auch aktuell fort – so nennen Lischka & Klingel die Evaluierung, ob der Einsatz eines ADM-Systems gegenüber anderen Lösungen tatsächlich zielführend ist, als eine der wichtigen Herausforderungen (Lischka & Klingel 2017, 14).

Menschen bauen noch stets parasoziale Beziehungen zu Computern auf. Im Verlauf seiner Forschung fiel Weizenbaum auf, dass Testpersonen während ihrer Gespräche mit ELIZA alleine gelassen werden wollten, als würden sie mit einem Menschen reden. Neuere Forschung zur Akzeptanz von Computern im therapeutischen Kontext zeigt, dass dieses Phänomen noch immer besteht; beispielsweise, dass Patienten sich eher einer App als einem menschlichen Therapeuten anvertrauen (Rost et al. 2017). Während Weizenbaum noch den Einsatz von Computern im Vietnamkrieg nutzen musste um zu zeigen, wie Menschen Verantwortung an Systeme abgeben die sie selbst nicht verstehen, ist dieses Phänomen mittlerweile in vielen anderen Lebensbereichen zu finden. Wie alltäglich die Akzeptanz von der vermeintlichen Autorität des Computers ist zeigen beispielsweise digitale Schwangerschaftstest, die auf einem Display anzeigen, ob der Test positiv oder negativ ist. Dieser digitale Test verspricht eine höhere, da computergestützte,

Genauigkeit eine Schwangerschaft zu erkennen. In Wirklichkeit passiert der eigentliche Test immer noch auf einem Papierstreifen der durch einen Lichtsensor ausgelesen wird. Der Sensor ist mit einem Prozessor verbunden der auf Basis der Sensordaten ein positives oder negatives Ergebnis berechnet und anzeigt. Dieser digitale Schwangerschaftstest unterscheidet sich damit im Kern nicht von herkömmlichen, analogen Tests die mit dem Auge durchgeführt werden müssen. Der bedeutende Unterschied ist jedoch, dass der Akt des Interpretierens und damit die Verantwortung für die Richtigkeit des Tests an einen Computer übergeben wurde (BBC 2020).

Seitdem Weizenbaum “Computer Power and Human Reason” geschrieben hat, gab es viele bedeutende Entwicklungen in der KI-Forschung und -Anwendung. ADM-Systeme werden im Jahr 2020 in vielen Bereichen des gesellschaftlichen Lebens eingesetzt und sind so alltäglich geworden, das sie kaum eine Besonderheit mehr darstellen. Die technischen Möglichkeiten von KI-Anwendungen haben sich insbesondere in den letzten 20 Jahren rasant entwickelt, angetrieben durch leistungsfähigere Rechner und der Verfügbarkeit großer Mengen von Daten. Neuartige Methoden wie neuronale Netzwerke ermöglichen es einerseits, Computer selbständig von Datensets lernen zu lassen, führen aber gleichzeitig dazu, dass Resultate als Black-Box-Modelle komplett intransparent sind. Eine Idee von Weizenbaum steht in seinen Ausführungen zentral: die Unterscheidung zwischen Entscheiden (“deciding”) und Wählen (“choosing”). Computer können nach fest programmierten Regeln Entscheidungen treffen aber nur Menschen haben eine echte Kapazität, ihre nächste Handlung auszuwählen. Dieser Punkt wird in aktuellen Diskussionen oft vernachlässigt. Gleichzeitig wird Menschen die Möglichkeit der Wahl genommen, wenn die Entscheidungen der ADM-Systeme nicht nachvollziehbar sind, auch wenn sie diese Systeme trotzdem weiter nutzen. Auch dieser Punkt war Weizenbaum bereits im Jahr 1976 bewusst; er verstand, dass das wahre Problem eine Illusion des Entscheidens ist: ”[...] But the widely believed picture of managers typing questions of the form “What shall we do now?” into their computers and then waiting for their computers to “decide” is largely wrong. What is happening instead is that people have turned the processing of information on which decisions must be based over to enormously complex computer systems. They have, with few exceptions, reserved for themselves the right to make decisions based on the outcome of such computing processes. People are thus able to maintain the illusion, and it is often just that, that they are after all the decisionmakers.

But, as we shall argue, a computing system that permits the asking of only certain kinds of questions, that accepts only certain kinds of ‘data’, and that cannot even in principle be understood by those who rely on it, such a computing system has effectively closed many doors that were open before it was installed.” (Weizenbaum 1976, 38).

Es ist also nicht der spezifische Akt, die Entscheidung eines Computers zu akzeptieren und zu übernehmen, sondern vielmehr die Einengung der Wahlmöglichkeiten durch das Medium Computer. Aus dieser Betrachtungsweise heraus erscheint es außerdem fragwürdig, inwiefern Maßnahmen wie Erklärbarkeit von Algorithmen oder rechtliche Ansprüche gegenüber algorithmischen Entscheidungen Weizenbaum positiv stimmen würden. Abgesehen von konkreten Problemen, beispielsweise dass Rechtsansprüche erst nach einer falschen Klassifizierung angestrengt werden können, ändern diese Maßnahmen nichts daran, dass diese Klassifizierungen überhaupt an ein Computersystem übergeben wurden, dass von vornherein die Möglichkeiten eingeschränkt hat. Der Problem ist also nicht ein schlecht funktionierender Algorithmus, das Problem ist der Einsatz des Algorithmus selbst. Dieser Gedanke findet sich auch in modernen Diskussionen wieder. So beschreibt Mark Bishop (2020) die gängige Meinung zu KI-Problemen als Kategorienfehler. Er weist darauf hin, dass Probleme mit KI-Anwendungen im öffentlichen Diskurs häufig als Rechenproblem dargestellt werden; der Algorithmus müsste nur schneller werden, oder optimiert werden, oder mit besseren Daten versorgt werden und diese Probleme wären aus dem Weg geräumt. Eine Gegenposition dazu, die das eigentliche Problem zumindest im Ansatz versteht, ist, dass diese Probleme nicht durch das bessere Erkennen von statistischen Mustern gelöst werden können, sondern dass das Maschinen kausale Zusammenhänge erkennen können müssen um diese Probleme zu vermeiden (Marcus und Davis 2019). Doch auch diese Position verkennt das Problem, dass Computer grundsätzlich, als Medium, nicht in der Lage sind Kausalität zu verstehen. Bishop (2020) beschreibt, dass auch dieser Ansatz falsch ist; zwar basieren tatsächlich einige Probleme mit KI-Anwendungen darauf, dass selbst komplizierte Black-Box-Anwendungen im Kern “nur” statistische Mustererkennung betreiben. Die Antwort darauf kann aber nicht sein, ihnen Kausalität beizubringen, denn Computer können grundsätzlich keine Kausalität verstehen wie Menschen sie verstehen. Da Computer, beziehungsweise “Computation”, kein Verständnis aufbauen kann, echte mathematische Einsichten haben kann und keine Sinneseindrücke (“raw sensation”)

verarbeiten kann, können sie niemals ein Verständnis von Problemen aufbauen, dass einem menschlichen Verständnis gleicht. Diese Limitierungen bestehen grundsätzlich im Medium Computer unabhängig vom spezifisch eingesetzten Algorithmus oder Machine-Learning-Methode (Bishop 2020, 32). Dieses fundamentale Problem beschreibt Bishop als “humanity gap”: “No matter how sophisticated the computation is, how fast the CPU is or how great the storage of the computing machine is, there remains an unbridgeable gap (a ‘humanity gap’) between the engineered problem solving ability of machine and the general problem solving ability of man” (Bishop 2020, 33). Auch in Weizenbaums Schriften lässt sich eine definitive Absolutheit in der Position zum Einsatz von KI finden: es gibt Bereiche, in denen KI-Technologien grundsätzlich nicht genutzt werden sollten. Diesem Gedanken liegt zugrunde, dass das Problem von Künstlicher Intelligenz nicht eine Frage von Schnelligkeit oder Verarbeitungskapazität ist, sondern Computer, aufgrund ihrer Eigenschaften als Medium, kategorisch ungeeignet sind, menschliche Belange zu beeinflussen. Diese Gedanken formulierte Weizenbaum bereits in den 1970ern als KI-Systeme noch weit entfernt waren von den Möglichkeiten moderner Systeme und zeigt, dass seine Kritik nicht nur im Jahr 2020 noch relevant ist, sondern dass er bereits in den Urzeiten der KI fundamentale, unlösbare Probleme erkannt hat.

Ohne dies Erkenntnis führt der Einsatz von ADM-Systemen und deren unkritische Akzeptanz zu einer von Algorithmen kontrollierten Gesellschaft. KI-Kritiker wie Pasquale (2015) argumentieren, dass das Ziel nicht nur Transparenz sein sollte, sondern eine Gesellschaft in der der Einsatz von Algorithmen komplett nachvollziehbar ist (vgl. Pasquale 2015, 189-218). Pasquale betont auch, ähnlich wie Weizenbaum, dass der Einsatz von Computer-Systemen eine vermeintliche Autorität aufbaut, die aber gleichzeitig die tatsächliche Effekte ihres Einsatzes verbergen und eine abstrakte Trennung aufbauen: “we should never lose sight of the fact that the numbers on their computer terminals have real effects, deciding who gets funded and found, and who is left discredited or obscure” (Pasquale 2015, 215).

Andere Kritiker beschreiben diese Form von Gesellschaft als “Algocracy” (Algokratie), die Herrschaft der Algorithmen und haben eine deutlich pessimistische Sicht, die aus der Einsicht entstanden sind, dass KI-Algorithmen fundamentale Probleme haben. Danaher

(vgl. 2016) beschreibt die “Algocracy” als einen gesellschaftlichen Zustand, in dem viele Entscheidungsprozesse im öffentlichen Bereich an Computersysteme abgegeben wurden und menschliche Teilnahme an öffentlichen Prozessen so eingeschränkt wird (Danaher 2016, 3). Die Gefahr durch diese “Algocracy” stellt nach Danaher unabhängig von spezifischen Problemlösungen durch Transparenz und Datenschutz eine Gefahr für die Legitimität von gesellschaftlichen Entscheidungen dar, da sie prinzipiell undurchsichtig sind. Als mögliche Lösung Antwort darauf schlägt Danaher vor, dass bei algorithmischen Entscheidungen immer Menschen eine letzte Prüfung vornehmen. Doch auch Danaher argumentiert, dass dies Problem nicht grundsätzlich löst und zieht eine pessimistische Bilanz: “ Furthermore, the growth of algocratic systems combined with the ways in which such system become woven into ever more complex algorithmic ecosystems, may be such as to push them beyond the control and understanding of their human creators. [...] In short, we may be on the cusp of creating a governance system which severely constrains and limits the opportunities for human engagement, without any readily available solution” (Danaher 2016, 34). Auch dieses Fazit deckt sich mit der Einschätzung von Weizenbaum, inwiefern eine von Algorithmen kontrollierte Gesellschaft vermeidbar ist oder nicht. Für Weizenbaum selbst ergibt sich noch nicht einmal die Frage, ob eine KI-gesteuerte Gesellschaft gewollt ist oder nicht: “There is not the slightest hint of a question as to whether we want this future. It is simply coming. We are helpless in the face of a tide that will, for no reason at all, not be stemmed. There is no turning back. Even the question is not worth discussing” (Weizenbaum 1976, 242).

Abschließend lässt sich die Forschungsfrage nach der heutigen Relevanz von Weizenbaums Kritik damit beantworten, dass viele von Weizenbaums Punkten auch fast 50 Jahre nach Erscheinen von “Computer Power and Human Reason” höchst relevant sind. Weizenbaums zentraler Punkt, der auch in aktuellen Diskussionen zu kurz kommt, ist sein Fokus auf den Computer als Medium und nicht auf seinen Inhalt, also Algorithmen und Anwendungen. Erst dadurch wird deutlich, dass bestimmte Probleme mit der Anwendung von KI, beispielsweise in Form von ADM-Systemen, fundamental sind und deswegen nicht gelöst werden können. Auch seine Einsichten zur Mensch-Maschinen-Interaktion – dass Menschen schnell parasoziale Interaktionen mit Computern eingehen, dass eine Diffusion von Verantwortung mit dem Einsatz von Computern einhergeht und dass die vermeintliche Autorität des Computers nur eine

Illusion ist, sind noch immer höchst relevant. Weizenbaums Schriften umfassen eine enorme Bandbreite von Themen und Ideen, die im Rahmen dieser Arbeit nicht oder nur unvollständig besprochen werden konnten, aber interessante fortführende Forschungsansätze darstellen. In seinen Schriften wird Weizenbaums sehr europäische Perspektive deutlich, die teilweise in starkem Kontrast zu dem Pragmatismus seiner angelsächsischen Kollegen steht, die seine Bedenken zu KI nicht teilen. Es wäre interessant, die Hintergründe hiervon zu erforschen und gegenüberzustellen.

6 Quellen und Referenzen

Absatzwirtschaft (2019): *Merkel: "KI made in Germany" als Gütezeichen*, https://www.absatzwirtschaft.de/merkel-ki-made-in-germany-als-guetezeichen-159985/ (Zugegriffen am 05.06.2020).

Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., & Rudin, C. (2017): *Learning certifiably optimal rule lists for categorical data*, in: The Journal of Machine Learning Research, 18(1), 8753-8830.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016): *Machine bias.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Zugegriffen am 10.02.2020).

Babe, Robert E. (2000): *The Communication Thought of Harold Adams Innis*, in Canadian Communication Thought: Ten Foundational Writers, Toronto: University of Toronto Press, S. 51-88.

Barocas, S. und Selbst, A. (2016): *Big Data's Disparate Impact*, in: California Law Review, 104(1), S. 671–729.

Barry-Jester, A.M., Casselman, B., & Goldstein, D. (2015): *The new science of sentencing*, https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing (Zugegriffen am 23.07.2020).

Bassett, Caroline (2019): *The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present*, in: AI & Soc 34, S. 803-812.

Bathaee, Y. (2017): *The artificial intelligence black box and the failure of intent and causation*, in: Harvard Journal of Law & Technology, 31, 889.

BBC (2020): *The surprising secret hidden in a pregnancy test*, https://www.bbc.com/news/technology-54025997 (Zugegrifen am 05.09.2020)

ben-Aaron, Diana (1985): *Weizenbaum examines computers and society*, http://tech.mit.edu/V105/N16/weisen.16n.html (Zugegriffen am 25.02.2020).

Block, Ned (1981): *Psychologism and Behaviorism, in The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, hrsg. von S. Shieber, Cambridge: MIT Press, S. 229-266.

Brinker et al. (2019): *Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task*, in: European Journal of Cancer, Volume 113, S. 47-54.

Buhrmester, V., Münch, D. & Arens, M. (2019): *Analysis of explainers of black box deep neural networks for computer vision: A survey*, in: arXiv preprint arXiv:1911.12116.

Carey, James (1967): *Harold Adams Innis and Marshall McLuhan*, in The Antioch Review, 27(1), S. 5-39.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Luwig, J., & Sendhil Mullainathan (2016): *Productivity and Selection of Human Capital with Machine Learning*, in: American Economic Review, 106 (5), S. 124-27.

Chouldechova, A. (2017): *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, in: Big data, 5(2), S. 153-163.

Danaher, John (2016): *The Threat of Algocracy: Reality, Resistance and Accommodation*, in: Philosophy & Technology, January 2016.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012): *Fairness through awareness*, in: Proceedings of the 3rd innovations in theoretical computer science conference, S. 214-226.

Gunning, D. (2017): *Explainable Artificial Intelligence (XAI)*, https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (Zugegriffen am 10.05.2020)

Hersh, Seymour M. (1973): *Senators Are Told U.S. Bombed Cambodia Secretly After Invasion in 1970*, https://www.nytimes.com/1973/08/08/archives/senators-are-told-us-bombed-cambodia-secretly-after-invasion-in.html (Zugegriffen am 08.02.2020).

Innis, Harold A. (2007), *Empire and Communications*. Toronto: Dundurn Press.

Jones, KM., McCoy, C. (2019): *Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework*, in: Learning, Media and Technology 44(1), S. 52-63.

Khandani, A.E., Kim, A.J., & Lo, A.W. (2010): *Consumer credit risk models via machine-learning algorithms*, in: Journal of Banking and Finance, 34, S. 2767–2787.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015): *Prediction Policy Problems*, in: American Economic Review, 105(5), S. 491-495.

LeCun, Y., Bengio, Y. & Hinton, G. (2015): *Deep learning*, in: Nature, 521, S. 436-444.

Lischka, K., & Klingel, A. (2017): *Wenn Maschinen Menschen bewerten*. Bertelsmann Stiftung: Arbeitspapier.

LKA Nordrhein-Westfalen (2018): *Projekt SKALA (Predictive Policing in NRW) - Ergebnisse,*
https://lka.polizei.nrw/artikel/projekt-skala-predictive-policing-in-nrw-ergebnisse
(Zugegriffen am 05.06.2020).

Long, M. (1985): *Turncoat of the computer revolution: An Interview with Joseph Weizenbaum*, in: New Age (December), 47-51, S. 76-78.

Marcus, G., and E. Davis (2019): *How to build artificial intelligence we can trust*, https://www.nytimes.com/2019/09/06/opinion/ai-explainability.html (Zugegriffen am 05.05.2020).

Markoff, John (2008): *Joseph Weizenbaum, Famed Programmer, Is Dead at 85*, https://www.nytimes.com/2008/03/13/world/europe/13weizenbaum.html (Zugegriffen am 10.03.2020).

Mayor, Adrienne (2018): *What Pandora's Box tells us about AI*, https://www.weforum.org/agenda/2018/10/an-ai-wake-up-call-from-ancient-greece/ (Zugegriffen am 07.07.2020).

McLuhan, Marshall (1994), *Understanding Media: The extensions of man*. Cambridge: MIT Press.

McLuhan, Marshall und Fiore, Quentin (1968), *War and Peace in the Global Village*. New York: Bantam Books.

Moor, James (2006): The Dartmouth College Artificial I*ntelligence Conference: The Next Fifty Years*, in: AI Magazine, 27(4), S. 87.

New York Times (1973): *...Admiral and Computer*, https://www.nytimes.com/1973/08/14/archives/admiral-and-computer.html (Zugegriffen am 08.02.2020)

Pasquale, Frank (2015): *The Black Box Society: The Secret Algorithms That Control Money and Information.* Cambridge: Harvard University Press.

Playboy Magazine (1969): *The Playboy Interview: Marshall McLuhan*, https://web.cs.ucdavis.edu/~rogaway/classes/188/spring07/mcluhan.pdf (Zugegriffen am 10.03.2020).

Ribeiro, M.T., Singh, S. & Guestrin, C. (2016): *Why should I trust you?: Explaining the predictions of any classifier*, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, S. 1142.

Rost, T., Stein, J., Löbner, M., Kersting, A., Luck-Sikorski, C., & Riedel-Heller, S. G. (2017): *User acceptance of computerized cognitive behavioral therapy for depression: systematic review*, in Journal of Medical Internet research, 19(9), E309.

Samek W., Müller KR. (2019): *Towards Explainable Artificial Intelligence. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, in: Lecture Notes in Computer Science, Vol 11700.

Sample, Ian (2017): *Computer says no: why making AIs fair, accountable and transparent is crucial*, https://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial (Abgerufen am 18.04.2020)

Searle, John (1980): *Minds, Brains, and Programs*, in: Behavioral and Brain Sciences 3, S. 417-424.

Tinnefeld, Marie-Theres (2019): *Künstliche Intelligenz - ein (digitales) Glasperlenspiel?*, https://rsw.beck.de/cms/?toc=ZD.20&docid=418857 (Zugegriffen am 10.07.2020).

Tolan, S. (2019): *Fair and unbiased algorithmic decision making: current state and future challenges*, in: arXiv preprint arXiv:1901.04730.

Turing, Alan M. (1950): *Computing Machinery and Intelligence,* in: Mind, Volume LIX, Issue 236, S. 433-460.

The Economist (2019): *Artificial intelligence is changing every aspect of war*, https://www.economist.com/science-and-technology/2019/09/07/artificial-intelligence-is-changing-every-aspect-of-war (Zugegriffen am 07.06.2020).

Tversky, A., & Kahneman, D. (1974): J*udgment under uncertainty: Heuristics and biases*, in: Science, 185 (4157), S. 1124-1131.

Wachter, S., Mittelstadt, B., & Russell, C. (2017): *Counterfactual explanations without opening the black box: Automated decisions and the GDPR*, in: Harvard Journal of Law & Technology, 31, S. 841f.

Wallace, Richard (2008): *The Anatomy of A.L.I.C.E*, in: Parsing the Turing: Philosophical and Methodological Issues in the Quest for the Thinking Computer, hrsg. von R. Epstein, G. Roberts & G. Beber, Heidelberg: Springer, S. 181-210.

Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013): *Learning to detect patterns of crime, in: Machine learning and knowledge discovery in databases*, CML PKDD 2013: Lecture Notes in Computer Science, vol 8190, S. 515–530.

Weizenbaum, Joseph (1966): *ELIZA—a computer program for the study of natural language communication between man and machine*, in: Communications of the ACM, 9(1), S. 36-45.

Weizenbaum, Joseph (1976), *Computer power and human reason*. San Francisco: Freeman.

Welchering, Peter (2018): *Wenn Computer über Menschen entscheiden*, https://www.swr.de/swr2/wissen/neue-regel-fuer-verbraucher-scoring,article-swr-13274.html (Zugegriffen am 08.07.2020).

Zweig, K. A., & Krafft, T. D. (2018): *Fairness und Qualität algorithmischer Entscheidungen*. In R. Mohabbat Kar, B. E. P. Thapa, & P. Parycek (Hrsg.), (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft, S. 204-227. Berlin: Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT).