Einleitung

Roger Häußling, Claudius Härpfer, Marco Schmitt

1. Künstliche Intelligenz – ein Thema in aller Munde!

Die Oral-B Genius X[1] ist eine elektrische Zahnbürste, die nicht nur dank mechanischer Rotation des runden Bürstenkopfes das händisch-menschliche Zähneputzen unterstützt und somit laut Herstelleraussage bis zu 100 % mehr Plaque entfernt als herkömmliche Zahnbürsten. Sie weist darüber hinaus die das Gerät bedienenden Menschen dank künstlicher Intelligenz auf Fehler beim Putzen hin. War man bei Vorgängermodellen noch darauf angewiesen, für ähnliche Funktionen stillzustehen und sich beim Putzen mit der Smartphone-Kamera in den Mund filmen zu lassen, so erfasst das 2019 erschienene Modell nun die Putzbewegungen mit der eingebauten Sensorik der Zahnbürste und errechnet über die mit Bluetooth verbundene Smartphone-Anwendung aus den Bewegungsmustern Putzstile, die dazu dienen, den Nutzer in Echtzeit auf Optimierungspotentiale beim Putzen hinzuweisen. Dieses Beispiel ist insbesondere in seiner vermeintlichen Trivialität fernab großer Erwartungen in Richtung einer starken, menschenähnlichen KI interessant (vgl. z.B. Zweig 2019: 267–280). Künstliche Intelligenz wird in einem abgesteckten Rahmen für eine spezifische Anwendung genutzt, um konventionelle Technik zu ergänzen und so schnellere oder bessere Ergebnisse zu erzielen, als dies rein mit konventionellen Mitteln möglich war. Diese Zahnbürste mit Zusatzfeature ist typisch für eine neue Art von alltäglichen Anwendungen und zeigt, wie verschiedene Faktoren einer soziotechnischen Innovation zusammenspielen. Bei genauem Hinsehen ist die (schwache) künstliche Intelligenz natürlich nicht in der Zahnbürste verortet, denn diese bietet mit der Sensorik lediglich einen Teil der Mensch-Maschine-Schnittstelle. Die Verarbeitung der Daten erfolgt über

1 https://www.oralb.de/de-de/produktkollektionen/elektrische-zahnbuersten/genius, abgerufen am 01.03.2023.

die gekoppelte Anwendung auf dem Smartphone, die wiederum auf einen großen Datenbestand zurückgreift, um Muster identifizieren zu können, mit denen typische Putzfehler korrelieren und diese dann auszugeben. Grundbedingung hierfür ist natürlich nicht nur, dass der nutzende Mensch bereit ist, Daten über sein tägliches Zahnputzverhalten mit dem Hersteller der Bürste zu teilen, sondern auch, dass er sein Smartphone zum Zähneputzen mit ins Badezimmer nimmt. Technische Entwicklung und soziale Konvention gehen hier Hand in Hand. Dass nun die Bürste – die also Erwachsenen nebenbei dabei helfen soll, ihr Zahnputzverhalten zu Optimieren – als intelligent beworben wird, spricht für die breite Akzeptanz und Verbreitung, die künstliche Intelligenz in den letzten Jahren im Alltag erlangt hat.

Es gäbe fraglos eine ganze Reihe anderer Beispiele. Fast 70 Jahre nach Schöpfung des Begriffs künstliche Intelligenz Mitte der 1950er Jahre (vgl. McCarthy/Minsky/Rochester/Shannon 2006) – für einen Abriss zur Geschichte der KI siehe Zweck/Werner in diesem Band – hat die Entwicklung, auch dank Datafizierung und Digitalisierung, dahingehend Fahrt aufgenommen, als dass unser Alltag mit Anwendungen durchdrungen ist, die sich künstlicher Intelligenz bedienen. Viele davon werden – wie die Zahnbürste – ohne große Kontroversen am Markt etabliert. Soziotechnische Meilensteine wie der im November 2022 veröffentlichte Textgenerator ChatGPT[2] hingegen sorgen für breite Diskussionen bis in die Tagespresse hinein und verdeutlichen damit die gesellschaftliche Relevanz des Themas. Hierbei hat es das Unternehmen OpenAI geschafft, einen Chatbot zu entwickeln, der auf Basis eines Sprachmodells und eines großen Datenpools in der Lage ist, nicht nur eine normale Konversation in bisher unerreichter Authentizität zu führen, sondern – wenn man ihn darum bittet – auch Hausarbeiten, Übersetzungen, Lyrik oder Programmcode zu schreiben. Er kann also im Gegensatz zur Zahnbürste eine Vielzahl von Aufgaben ausführen, die das Intelligenzniveau konventioneller Technik überschreiten. Diese Entwicklung hat das Potential das Arbeiten mit Texten massiv zu verändern und macht auch vor der kreativen Seite des Arbeitens nicht halt. Die Grenzen dieser Entwicklung sind derzeit noch nicht abzusehen. Was die kurz nach Veröffentlichung des Tools entflammte Diskussion um ChatGPT aber zeigt, ist, dass schnell rechtliche und ethische Fragen im Fokus stehen (vgl. z.B. Bleher/Braun 2023; Nehlsen/Fleck 2023), während die langfristig relevante soziologische Perspektive ins Hintertreffen gerät.

2 https://openai.com/chatgpt, abgerufen am 01.03.2023.

GPT steht für »Generative Pretrained Transformer« und basiert auf einem Deep Learning Verfahren mittels eines künstlichen neuronalen Netzes – also im Grunde auf nichts Neuem. Insofern ist erklärungsbedürftig, warum ChatGPT so viel Aufmerksamkeit erfahren hat, wo es doch haufenweise andere Anwendungen von Deep Learning Verfahren gibt. Von Nutzer:innenseite springen dabei zwei Dinge ins Auge: Die Niederschwelligkeit seiner Nutzung und die Qualität seiner Ergebnisse. Sowohl die Registrierung für eine kostenfreie Nutzung dieser Künstlichen Intelligenz, als auch die Handhabung ihrer Eingabemaske geschehen im Handumdrehen, ohne dass jeweils spezifische Kenntnisse notwendig oder besondere Hürden zu nehmen wären. Die Frage der Qualität ist engstens – wie im vorliegenden Band mehrfach herauszuarbeiten sein wird – mit der Quantität und Qualität der Lerndaten verknüpft. Hier hat OpenAI neue Maßstäbe gesetzt, ohne jedoch irgendwelche innovativen neuen Verfahren zu etablieren, die vorher bei anderen KI-Verfahren nicht zur Anwendung gekommen wären. Vielmehr wurden etablierte Verfahren des künstlichen Lernens mehrstufig zur Anwendung gebracht, sodass sich schon die Frage stellt, ob wir es hier wirklich mit einer neuen Stufe künstlicher Intelligenz zu tun haben.

Eine Besonderheit weist GPT allerdings doch auf, die deren Neuheitsanspruch unterstreichen könnte: Während bisherige KI-Verfahren durch eine Fokussierung auf ein konkretes Problem und dabei auf einen bestimmten Analysegegenstand gekennzeichnet waren, ist GPT breiter in Bezug auf Probleme und Analysegegenstände angelegt. Dies wird besonders bei GPT-4 deutlich, insofern nicht nur Text, sondern auch Bilder als Eingabe erfolgen können. Aus einem Bild als Vorlage kann dann GPT-4 zum Beispiel einen Code für eine Homepage generieren. Bereits ChatGPT-3 kann sowohl lyrische Texte in einem gewünschten Stil, wissenschaftliche Abhandlungen zu verschiedensten Themen, politische Reden aus unterschiedlichen parteipolitischen Positionen heraus oder Quellcode für verschiedene Anwendungen generieren – je nach dem, worum man die Künstliche Intelligenz konkret bittet. Genau hierin unterscheidet sich GPT von bisherigen KI-Verfahren!

Das erklärte Ziel der Macher:innen von OpenAI ist es, an der Entwicklung einer so genannten AGI, Artificial General Intelligence zu arbeiten – einer starken Künstlichen Intelligenz also, die für alle möglichen Aufgaben zum Einsatz gebracht werden kann. Auf technischer Ebene bleibt allerdings unklar, wie genau GPT-4 Schritte in diese Richtung vollzieht. Vormals waren Bilderkennung und Textmining zwei getrennte Angelegenheiten, für die jeweils eigene Learner zum Einsatz kamen. Wie nun Text- und Bilddaten zusammengeführt wer-

den, bleibt ein streng gehütetes Geheimnis von OpenAI, sodass der eigentliche Neuheitswert nicht wirklich unter die Lupe genommen werden kann.

Bleibt noch eine weitere Besonderheit von GPT: Wird die Datenlage dünn, dann fängt GPT zu halluzinieren an, sprich: die KI erfindet kontextbezogen Dinge hinzu – und zwar oftmals so gekonnt, dass es für Nicht-Sachkundige nicht gleich erkennbar ist. Das erzeugt auf der Nutzer:innenseite im Umgang mit der KI generell Unsicherheiten, da man bei den von ihr angebotenen Ergebnissen nie genau wissen kann, ob Halluzinationsanteile enthalten sind. Im Grunde äußert sich darin aber auch eine besondere Qualität der Halluzinationen. Sie sind oftmals doch so gut, dass sie für uns als Fakten erscheinen. Diese ›kreative‹ Leistung von GPT wird bislang nur selten herausgestellt. Immerhin gelingt es der KI, menschlichen Nutzer:innen mitunter täuschend echte Phantasmagorien aufzutischen. Eine solche Münchhausen-Qualität hätte man vor ein paar Jahren einem künstlichen neuronalen Netz nicht zugetraut. Zu Recht wird auf die Gefahr einer Potenzierung von »Fake News« durch solche Dienste hingewiesen, dabei aber auf die Kehrseite des Phänomens bislang wenig abgehoben: Sind diese halluzinatorischen Eigenschaften von GPT nicht auch gezielt nutzbar für spekulative Problemstellungen, um uns mit einer Vielfalt an sinnvoll erscheinenden Brückenargumenten bezüglich unserer Wissenslücken zu versorgen?

Vor diesem Hintergrund haben wir ChatGPT zu seiner technischen Verfasstheit, zu seiner Funktionsweise sowie zu Chancen und Risiken von Deep Learning im Allgemeinen befragt. Auch wollten wir schauen, ob sich die KI in die Enge treiben lässt, wie sie sich eine Soziologie des Deep Learnings vorstellt, und was sie von netzwerkforscherischen Zugängen zur KI hält. Die Ergebnisse dieses Interviews finden sich am Ende dieses Bandes.

Für eine intensive soziologische Auseinandersetzung mit der neueren Künstlichen Intelligenz fehlen bislang zum einen noch Möglichkeiten einer empirischen Auseinandersetzung mit dem Thema jenseits öffentlicher oder wissenschaftlicher Diskurse und zum anderen auch ein theoretischer Zugang der techniksoziologisch, wie sozialtheoretisch überzeugt. Diese Lücke möchte der Sammelband schließen. Es liegt nahe, einen solchen Zugang in der Netzwerktheorie zu suchen, da sie begrifflich wie methodisch auf das kalibriert ist, was auch (künstliche) neuronale Netzwerke ausmacht. Der Sammelband versucht entsprechend auf der Basis einer dezidiert relationalen Theorieperspektive und ausgewählter Anwendungen, soziologische Analyseangebote für das Phänomen einer, große Teile von Wirtschaft und Gesellschaft durchwirkenden, Künstlichen Intelligenz zu entwickeln. Hierbei erweist sich

die Theorie von Harrison White als besonders vielversprechend (systematisch hierzu Häußling/Schmitt in diesem Band). Mit der Ableitung zentraler Theoriekomponenten aus einer Vielzahl von empirischen Arbeiten aus der Netzwerkforschung und den grundlegenden Komponenten von Verteilung und Selbstähnlichkeiten bietet White auch eine dem Phänomen der neueren KI stark angenäherte Theoriesprache, da beide stark auf Mustererkennung setzen. Seine zentralen Begriffe Identität, Kontrolle, Netzwerkdomäne, Switching, Stil und Kontrollregime ermöglichen zudem aufgrund ihrer Abstraktheit ihre direkte Anwendbarkeit auf KI-Phänomene par excellence. Gerade weil KI im Kern netzwerkartig ist, kann Whites elaborierte Netzwerktheorie so fruchtbringend daran angeschlossen werden und neue Einsichten in das Phänomen KI liefern.

2. Zum Begriff der »neueren KI«

Unter dem Begriff der neueren KI verstehen wir in erster Linie Anwendungen, die auf Machine Learning (ML) und Deep Learning (DL) zurückgreifen. Die grundlegende Engineering-Idee dieser in den letzten Jahren aufgekommenen Anwendungen ist, nicht mehr alles programmieren zu wollen (in manchen Fällen, wie zum Beispiel bei autonomen Fahrzeugen, zu können), sondern stattdessen künstliche neuronale Netze darauf zu programmieren, Muster zu erkennen. Die dabei zum Einsatz kommenden Programme, die so genannten Learner, sind vergleichsweise schlicht und lassen sich – wie die oben erwähnte Zahnbürste – in vielen Fällen problemlos als Smartphone-Anwendung realisieren, komplex hingegen sind die Datenmengen, anhand derer KI trainiert wird. Dies ist möglich durch den erleichterten Zugang zu großen Datenmengen durch technische und soziale Entwicklungen und die Fortschritte im Bereich der Machine Learning-Verfahren, insbesondere auf der Basis neuronaler Netze.

Kern dieser Learning-Verfahren sind jene künstlichen neuronalen Netze (vgl. z.B. Rosengrün 2021: 24–29; Mainzer 2019: 104–124). Vom menschlichen Nervensystem inspiriert, werden komplexe Schaltungen aufgebaut, in denen Neuronen als Knoten ›feuern‹, also Verbindungen zu anderen Neuronen aufbauen, wenn sie aktiviert werden, weil ein bestimmter Grenzwert erreicht ist. Als Stellschrauben dienen zum einen die Gewichtung des Inputs und zum anderen die Höhe des Grenzwerts. Diese Netze sind so aufgebaut, dass es eine Eingabeschicht (Input Layer), in der Regel eine oder mehrere Zwischenschich-

ten (Hidden Layers) und eine Ausgabeschicht (Output Layer) gibt. Durch diese Struktur können große systematische Zusammenhänge modelliert werden, indem der Learner im Rahmen seiner Rechenleistung die möglichen Fälle für die einzelnen Neuronen und Schichten nacheinander durchspielt und so statistische Korrelationen erhält, die irgendwann mit Blick auf den weiteren Aufwand ein befriedigendes Ergebnis liefern.

In der Lernpraxis unterscheidet man zwischen drei grundlegenden Verfahren, um unterschiedliche Funktionen zu erfüllen. Das gängigste Verfahren ist das sogenannte Supervised Learning, bei dem der Mensch dem Learner einen (vergleichsweise schmalen) Trainingsdatensatz zur Verfügung stellt, in dem der Learner Input-Daten und Output-Daten vergleichen kann, um die Zusammenhänge direkt zu sehen. Ist das künstliche neuronale Netz auf Basis dieser Lerndaten optimiert, kann es operativ eingesetzt werden. Wenn diese Möglichkeit des Blicks auf vorab vergangene Daten nicht gegeben ist, also kein Feedback vorliegt, spricht man von Unsupervised Learning. Hier muss sich der Learner die Daten selbst annotieren, indem er Muster findet und Kategorien bildet, also auf Basis der Input-Daten, seine Output-Daten schafft. In diesem Fall ist natürlich ein größerer Datenpool nötig. Bei der dritten Form, dem Reinforced Learning, hingegen, werden die Output-Daten im Prozess evaluiert, um die Mustererkennung zu optimieren. Die drei Verfahren verfolgen unterschiedliche Zwecke und werden in der Praxis gerne auch kombiniert.

Die auf Basis dieser Verfahren entwickelten Anwendungen sind aus unserem Alltag nicht mehr wegzudenken und haben dementsprechend eine nicht zu unterschätzende Wirkung auf unser soziales Leben. Daher ist es unabdingbar, diese Phänomene einer adäquaten soziologischen Betrachtung zu unterziehen.

3. Neuere KI in der sozialwissenschaftlichen Diskussion

Während Machine Learning und dessen (potentielle) Nutzbarmachung für die Sozialwissenschaften immer wieder diskutiert werden (vgl. Molina/Garip 2019; Mökander/Schroeder 2021; Heiberger 2022), scheint uns zunächst insbesondere das Diskursfeld der so genannten Critical Code Studies ein fruchtbarer Nährboden für sozialwissenschaftliche Reflektionen zu sein. Hierin lassen sich, wie wir anhand einiger ausgewählter Studien detaillierter zeigen, fünf zentrale Topoi ausmachen, um die die gegenwärtigen sozialwissenschaftlichen Diskussionen rund um die neue Künstliche Intelligenz kreisen. Im

Einzelnen sind dies: (1) Fragen zu dem Wechselverhältnis zwischen Kontrolle und Machine Learning, (2) die durch den Einsatz künstlicher neuronaler Netze eingeläutete Abkehr von deterministischen bzw. kausalistischen Verfahren, (3) – damit eng verknüpft – die Hinwendung zu Realexperimenten durch die Tatsache, dass es sich um lernende Systeme handelt, (4) die Anzeichen für eine postdigitale Ära sowie (5) die Opakheit von Deep Learning-Verfahren. Diese Topoi werden im Folgenden in der gebotenen Kürze dargelegt.

Zum Wechselverhältnis von Kontrolle und Machine Learning

Einen wichtigen Beitrag zur Frage nach dem Wechselverhältnis von Kontrolle und Machine Learning hat Mackenzie (2017) geleistet. In Anlehnung an Deleuze spricht er von einer Assemblage, die Menschen gemeinsam mit den Machine Learnern bilden. Nur auf der Ebene der Assemblage sei das Phänomen des Lernens adäquat verstehbar. Es – mit anderen Worten – allein auf der Ebene der künstlichen neuronalen Netze ergründen zu wollen, verkennt, wie viel an den Vorgängen in der Trainingsphase dem menschlichen Zutun geschuldet ist. Selbst bei den Schematisierungsleistungen des Machine Learners sind für Mackenzie menschliche Ideen, Gedanken und Empfindungen eingegangen. Es wäre also falsch, anzunehmen, die beteiligten Technologien würden grundlegend anderen Prinzipien folgen als diejenigen der Menschen. Das Gegenteil ist der Fall: Der Machine Learner ist für Mackenzie durch und durch Ausdruck des menschlichen Willens, Wissen und damit Macht zu erlangen. Die Besonderheit dieser Technologien liege darin, dass sie in Bereiche vorzustoßen vermögen, die bislang für den Menschen verschlossen waren. In riesigen Datenmengen spürten sie Muster und Schemata auf und würden damit überhaupt erst komplexe Phänomene kontrollierbar machen, was zu neuen Formen der Akkumulation von Machtwissen führe.

Mackenzie vermutet sogar hinter den Lernprozessen des Machine Learners eine »Technologie des Selbst«; denn die Mustererkennungen und Schematisierungen dienten ja nur dazu, um sich noch besser an die Gegebenheiten der Umwelt anzupassen. Nichts Anderes seien gemäß Foucault »Technologien des Selbst« (1993), die im Rahmen einer Assemblage auch bei den Menschen, die sich auf Deep Learning einlassen, anzutreffen seien: Sich als Subjekt zum Objekt zu machen, Erwartungshaltungen des sozialen und gesellschaftlichen Umfelds so zu internalisieren, dass sie ununterscheidbar von ›ureigenen‹ Bedürfnissen und Strukturen werden, kurzum: sowohl als Subjekt als auch als Objekt von Machine Learning in Erscheinung zu treten. Bereits heu-

te ist für Mackenzie sichtbar, dass uns dabei die Machine Learner verändern. Ihre ubiquitären Verwendungsmöglichkeiten führten im Ergebnis dazu, dass es letztendlich ununterscheidbar werde, was auf unseren eigenen Gedanken fuße und was durch Machine Learning ›errechnet‹ sei. Aber genau hierin sei die Wirkmacht der »Technologie des Selbst« zu suchen: Die komplexer werdende Gesellschaft kann für Mackenzie nur gemeistert werden, wenn sich der Mensch auf die Machine Learner einlässt, was zwangsläufig zu einer Veränderung seiner selbst führt. Mackenzie spricht von einer Vorwärtsregelung, die in der Trainingsphase des Learners zu beobachten ist. Hierbei gehe es weniger um Verbesserung der Erfahrung, denn um das tentative Auffinden neuer Relationen, um drängende Realitäten verstehbar und bearbeitbar zu machen.

Dabei steht nach Mackenzie Machine Learning für ein komplett anderes Verständnis, wie Computeroperationen wirkungsvoll zu kontrollieren sind – nämlich jenseits der klassisch-deterministischen Programmierung. Denn Machine Learning sei nur im Horizont einer Krise der Kontrolle, die durch das exponentielle Wachstum des Internets der vergangenen Jahrzehnte verursacht worden sei, zu verstehen. Die operative Macht des Machine Learnings ergibt sich für Mackenzie aus dem Versprechen, im Wildwuchs der Datenströme und der digitalen Kommunikationsprozesse Kontrolle wiederzuerlangen. Dabei sei Machine Learning eine verschachtelte Form von Kalkulation, um zukunftsorientierte Entscheidungen zu ermöglichen.

Mackenzie fordert von den Sozialwissenschaften in diesem Zusammenhang, auf dem gleichen Abstraktionsniveau wie Algorithmen zu forschen, um die Vielfalt der Kategorisierungs- und Klassifikationsformen bei Machine Learning in den Blick zu bekommen. Denn die Differenz der Praktiken wirke sich in der im vorhergehenden Absatz beschriebenen Weise auf das Welt- und Selbstverständnis des Menschen aus. Erst die genaue Kenntnis dieser Praktiken ermögliche es, ihnen kritische Praktiken und Experimente entgegenzusetzen, um die Veränderbarkeit dieser abstrahierenden Prozesse, Kalkulationen und Automationen unter Beweis zu stellen und daran anschließend revolutionäre Institutionen aus Daten und Codes zu schaffen. Denn – gemäß Mackenzies sozialutopischer Vision – ist der Mensch als Element dieser soziotechnischen Assemblage stets in der Lage, das Kollektiv – sprich: die Assemblage selbst – zu verändern.

Ganz im Sinne dieses Assemblage-Gedankens verteilter Kontrollkonstellationen erkundet Engemann (2018) die Rolle des menschlichen Körpers bei ML. Er sei indexikalischer Zeichengeber, der vor allem sich bei der Erstellung der Trainingsdaten und dem sog. Labeling bemerkbar mache. Hier

fungiere der Mensch als Relais, der zwischen Datum, Label und Wirklichkeit einen Zusammenhang herstelle. Im haptischen Sinne erfolge dies durch millionenfaches Clickwork, also den menschlichen Handlungsvollzügen des Bestätigens via Mausklick. Erst dadurch entstehe der Konnex zwischen Zeichen und Referenten, der so bedeutsam für die Güte der Datensätze in Bezug auf die ML-Lernzwecke und damit für die Wirksamkeit des ML-Verfahrens in der Praxis seien. Zwar wird – so Engemann – fieberhaft an unsupervised-Verfahren geforscht, die Realität sieht aber nach wie vor gänzlich anders aus: Nur durch das menschliche Zutun, werden die Verfahren in die Lage versetzt, überhaupt so etwas wie Präzision zu entwickeln.

Abkehr von deterministischen bzw. kausalistischen Verfahren

Ein wichtiger Topos der sozialwissenschaftlichen Befassung mit Machine Learning- und Deep Learning-Verfahren bildet der Aspekt, dass sie weder deduktiv noch induktiv vorgehen. In diesem Sinn bildet etwa für Parisi (2018) Machine Learning eine grundlegend andere Herangehensweise als die bis dahin vorherrschenden algorithmischen Strategien der Erzeugung von Künstlicher Intelligenz. Denn konstitutiv für das ›neue‹ maschinelle Lernen sei, dass das Unvollständige, Unbestimmte, Undeterminierte und damit letztlich Unberechenbare eine konstitutive Rolle für die internen Lernvorgänge spielten. Im Kern geht es – gemäß Parisi – um abduktiv gewonnenes Wissen, wobei abduktive Verfahren der Erkenntnisproduktion nicht etwa im Laufe der Zeit zur Auflösung dieser Unbestimmtheit führt, sondern im Gegenteil zu dessen Bewahrung. Statt wie früher durch Top-Down-Programmierung Ansatzpunkte für maschinelles Lernen zu suchen, seien es nunmehr automatisierte Systeme, die sich bei der »Trial-and-Error-basierten Datengewinnung durch schnelle, unbewußte und nicht-hierarchische Befehle von Entscheidungsprozessen auszeichnen« würden (vgl. ebd.: 99).

Diese nicht-deduktiven Verfahrensweisen öffneten das Tor zu Vorhersagen und Mustererkennungen. Diese Verfahrensweisen bildeten mit statistischen Kalkulationen zusammen eine Synthese, bei der die digitalen Prozeduren lernen, ihre Pfade durch die Datensuche zu verändern. Dabei werde gerade anhand von kontext-spezifischen Dateninhalten und dem Scharfstellen auf deren Besonderheiten sowie auf Überraschungsmomente gelernt. Für Parisi geht es um die Generierung hypothetischer Bedeutung, bei der die Undeterminiertheit ein aktiver Teil der Berechnung geworden ist. Damit lernten die Maschinen das Lernen selbst, als eine Unwissenheit bewahrende Aktivität, welche

weitere Lernkurven ermögliche. Denn die Lückenhaftigkeit der Berechnung eröffne die Welt der Zufälligkeit innerhalb der algorithmischen Abläufe und damit den Weg zu neuen Ideen, Verhalten Mustern und Regeln.

Bächle et al. (2018) verweisen hierbei auf den »Differentiable neural Computer« (DNC) von DeepMind/Google, dessen künstliches neuronales Netz Besonderheiten aufweist, was ihm die Kennzeichnung eines »externalisierten Gedächtnisses« eingebracht hat. So würden Informationen von DNC nicht eindeutig, sondern in Form einer Verteilung gespeichert, auch schwach ausgeprägte Ähnlichkeiten berücksichtigt, aufgrund dieser beiden Besonderheiten konkrete Problemstellungen sowie die Größe des dann faktisch in Aktion tretenden künstlichen neuronalen Netzes festlegen und Inhalte selbst nach ihrer Abspeicherung veränderbar seien – um nur die wichtigsten Besonderheiten zu nennen. Mit diesen Besonderheiten solle das technische System in die Lage versetzt werden, selbsttätig Abwägungen und Beurteilungen vornehmen zu können sowie Informationen kreativ für die Entwicklung eines Lösungspfades für konkrete Probleme zu entwickeln, sich diesen zu merken und für ähnlich gelagerte Problembearbeitungen generalisieren zu können. Darin äußert sich – so die Autor:innen – ein »diagrammatisches Denken« im Sinne Charles S. Pierce[3], der diesem Denken spezifische Transfer- und Übersetzungsfähigkeiten zuspricht.

Die Hinwendung zu Realexperimenten

Mit einer Veränderung des logischen Schließens geht auch eine Veränderung des Zugriffs auf die empirische Wirklichkeit einher. Unter dem Topos Realexperimente wird auf den Aspekt der neuen Künstlichen Intelligenz scharfgestellt, dass es sich um lernfähige und damit stets in Optimierungsschleifen befindliche Systeme handelt. Stilgoe (2018) wählt zur Exemplifizierung dieses Sachverhalts das autonome Fahrzeug Modell S von Tesla, welches im Mai 2016 einen tödlichen Unfall produziert hat, da es einen weißen Laster an einer

3 Das Diagramm bildet für Pierce eine besondere Klasse ikonischer Zeichen. Sie weisen nämlich – so Pierce – die nur ihnen zukommende Charakteristik auf, eine ›intelektuelle‹ Verbindung zwischen bildlichen und symbolischen Elementen herzustellen. Das, was sie repräsentieren, verdeutlichen sie mit anderen Worten bildlich, aber auch zeichenhaft. Insofern korrespondieren Diagramme für ihn mit dem abduktiven Denken, insofern von einer Beschäftigung mit einem Diagramm mehr gelernt werden kann, als in ihre Verfertigung eingeflossen ist (vgl. Peirce 1976: 47–54).

Straßenkreuzung übersehen hat. Ausgerechnet technische Unfälle ermöglichten, den Forschenden und Entwickelnden die Kontrolle über das Realexperiment zu entziehen und auf die mehr oder weniger impliziten Unsicherheiten scharfzustellen. Denn gerade, weil die neue Künstliche Intelligenz dort eingesetzt werde, wo klassische Programmierung scheitere – nämlich bei komplexen Aufgaben, bei denen es zu viele mögliche Situationen geben könne, als dass diese im klassischen Sinne programmiert werden könnten –, gerade deshalb befänden sich die Machine Learning-Verfahren in einem andauernden »work-in-progress«- bzw. »Beta«-Zustand. Ja, nur der Einsatz in realen Situationen ermöglichte derartigen Systemen, im Hinblick auf die komplexe Aufgabe, angemessen zu lernen.

Für Stilgoe führen die Begriffe »selbstfahrend« oder »autonom« in die Irre. Vielmehr handle es sich um Autos, die lernen würden, zu fahren – und dies mitten unter uns. Insofern sei es nicht verwunderlich, dass die Entwickler:innen von Tesla die Antwort auf den Crash in der Fortsetzung der Forschung an den ML-Verfahren sähen. Jedoch, so Stilgoe, bezieht Tesla damit gleichzeitig eine Position, der gemäß das Lernen dieser Systeme im privaten Hoheitsgebiet von Tesla liegt, was hochproblematisch ist. Vielmehr müsse das Lernen umfassender gedacht werden: im Sinne eines sozialen Lernens, das nicht nur die technischen Systeme betreffe, sodass sie sich besser in einer sozialen Welt zurechtfänden, sondern darüber hinaus wie alle Anspruchsgruppen im Hinblick auf ihren alltäglichen Umgang mit der neuen Technologie und wie die Gesellschaft als Ganzes im Sinne einer Governance neuer Technologien lernten. So gesehen, gäbe es weitere Optionen, produktiv an den Crash anzuknüpfen – etwa mit der Governance-Option, Konzerne wie Tesla dazu zu verpflichten, ihre Daten zu teilen, sodass nicht jedes Unternehmen selbst solche schmerzhaften Erfahrungen machen und Schäden produzieren müsse. Eine Technologieentwicklung, die auf ein so verstandenes soziales Lernen abhebt, ist Stilgoe zufolge resilienter und kann aus Krisen sogar gestärkt hervorgehen. Gerade Machine Learning biete hierzu eine hervorragende Möglichkeit, einen Take off des (sozialen) Lernens zu vollführen: denn prinzipiell brauche nur ein einziges Auto eine bis dahin für das System unbekannte Situation zu lernen, sodass alle lernfähigen Autos über Auto-zu-Auto-Kommunikation (Stichwort: Internet der Dinge) davon lernten. Entsprechend mündet Stilgoes Argumentation in der Forderung nach einem »social ML«, also lernende Systeme in ihre soziale Welt zurückzubetten und Technologieentwicklung als einen ganzheitlichen Prozess des sozialen Lernens zu reframen.

Anzeichen einer postdigitalen Ära

Mindestens ebenso oft wie auf die Betaisierung technischer Systeme wird in der sozialwissenschaftlichen Forschung auf dessen postdigitale Aspekte abgehoben. Prominent vertritt diese These Sudmann (2018), indem er auf das in Deep Learning-Verfahren bedeutsame Backpropagation hinweist, bei dem es im Kern um die Anpassung von Gewichten zwischen Verbindungen im betreffenden künstlichen neuronalen Netz geht. Dabei betont auch Sudmann (wie auch Parisi, s.o.) eine Buttom-up-Strategie in der neueren KI, also auf digitalem Wege ein Gehirn nachzumodellieren (im Unterschied zu der alten KI mit ihren Top-down-Strategien, wie dem Anlegen großer Wissensspeicher). Nunmehr würde die Nachbildung des biologischen Gehirns das vormals vorherrschende kognitivistische Verständnis verdrängen.

Der Fokus auf künstliche neuronale Netze stellt für Sudmann

> »in mindestens zweifacher Hinsicht ein Gegenmodell zur Funktionsweise digitaler Computer gemäß der seriell organisierten Von-Neumann-Architektur dar: (1) Die Gewichtung der Aktivität zwischen den Neuronen, d.h. die Stärke ihrer Verbindungen, wird bei neuronalen Netzen zumeist durch Fließkommazahlen, also quasi-analog, repräsentiert. (2) Die massenhaft miteinander verbundenen Neuronen feuern gemeinsam bzw. parallel, und bilden auf diese Weise ein komplexes emergentes System, das letztlich die Diskretheit der Elemente, aus denen es besteht, fundamental aufhebt.« (Ebd.: 66)

Entsprechend bilde die neue KI auch ein Gegenmodell zur Welt der klassischen Algorithmen, in der durch serielles Operieren innerhalb eines wohldefinierten Problemraums alle auftretenden Probleme logisch gelöst werden konnten. Demgegenüber steht die neue KI für ein konnektionistisches Paradigma: In Entsprechung zum grauen Rauschen im Gehirn liegt mit einem künstlichen neuronalen Netzwerk ein »Chaos« an (potentiellen) Verknüpfungen vor, innerhalb dessen Muster und Sinnzusammenhänge als emergente Verkopplungsphänomene erscheinen – so Sudmann. Bedeutung ergebe sich also aus spezifischen Systemzuständen des künstlichen neuronalen Netzes. Dieses Netz »ist demzufolge ein Unschärfesystem mit probabilistischen Resultaten, dessen Operationen eher als analog denn als digital zu beschreiben sind« (ebd.: 67). In letzter Konsequenz erwiesen sich dadurch auch die bisherigen Computer, die weitestgehend nach der Von-Neumann-Architektur aufgebaut sind, als unzu-

reichend, weil sie gerade auf das serielle Operieren spezialisiert seien. Demgegenüber stünden Technologien wie die Tensor Processing Units (TPUs), die Neurocomputer sowie die Quantencomputer für aussichtsreiche Forschungsaktivitäten, die den Weg zum ›analogen Rechnen‹ und damit zu einer postdigitalen Ära bahnten.

Opake Systeme

Schließlich besteht ein weiterer bedeutsamer Topos in der sozialwissenschaftlichen Forschung zur neuen Künstlichen Intelligenz in ihrem opaken, intransparenten Charakter. Gerade bei Deep Learning-Verfahren bleibt es in der Regel unklar, wie sie zu einem Ergebnis gelangen – und dies selbst dann, wenn man die Inputdaten genau kennt (was selbst schon oftmals nicht der Fall ist). Dies ist insofern für die Sozialwissenschaften von besonderer Relevanz, da diese Verfahren zum Beispiel für die Berechnung der Kreditwürdigkeit einer Person eingesetzt werden. D.h. der Einsatz dieser Verfahren zeitigt heute schon erhebliche soziale und gesellschaftliche Konsequenzen, ohne dass deren Ergebnisproduktion verstanden, nachvollziehbar und damit legitimierbar gemacht werden kann. Burrell (2016) differenziert dabei drei Formen der Undurchsichtigkeit dieser Verfahren:

1. Sie kann zum einen daraus resultieren, dass sie einer organisationalen oder staatlichen Geheimhaltung unterliegen. Man denke an NSA oder die großen Plattformkonzerne, deren Geschäftsmodell nicht zuletzt gerade darin besteht, immer ausgeklügeltere Big Data-Analyseverfahren zu entwickeln, um zum Beispiel an tiefere Schichten des menschlichen Verhaltens heranzukommen, um es zu manipulieren.
2. Die Undurchsichtigkeit von Deep Learning-Verfahren kann darüber hinaus auch schlicht daran liegen, dass wir nicht über den nötigen technischen Sachverstand verfügen, um ihre Operationsweise zu verstehen.
3. Undurchsichtigkeit kann aber auch dadurch entstehen, dass die durch diese Verfahren beschrittenen mathematischen Lösungswege in Big Data-Kontexten nichts mehr mit dem zu tun haben, wie menschliches Verstehen vonstattengeht. Sie folgen mit anderen Worten ganz anderen Anforderungen als diejenigen, die menschliche Einsichten und semantische Interpretationen ermöglichen.

Unter Rückgriff auf eine Forderung von Pasquale (2015) sieht Burrell die erste Form von Undurchsichtigkeit durchaus als auflösbar an: Denn dass Organisationen und staatliche Institutionen Geheimhaltung in der angesprochenen Form betreiben können, hat ja letztlich mit einem Vakuum an Regulation zu tun. Wären diese Organisationen gesetzlich dazu verpflichtet, ihre Datenanalyseverfahren offenzulegen – etwa durch Etablierung staatlicher Prüfstellen und Auditoren, welche zu weitgehende Manipulationen oder Ausspionierungen unseres Verhaltens ahnden können.

Auch die zweite Form von Undurchsichtigkeit lässt sich gemäß Burrell gesellschaftlich abmildern, wenn nicht gar heilen: Zum einen könnte die Ausbildung und Erziehung einen wesentlich stärker ausgeprägten Fokus auf Computer-Skills, insbesondere Programmierfähigkeiten, für alle legen. Zum anderen könnte man die Programmierer:innen dazu verpflichten, Entschlüsselungshilfen für ihre Codes mitzuliefern, sodass sie nicht nur für Computer lesbar, sondern auch für Menschen nachvollziehbar werden, was unter dem Begriff explainable AI verhandelt wird.

Von besonderem forscherischem Interesse ist jedoch die dritte Form von Undurchsichtigkeit für Burrell: Denn selbst wenn die Datensätze, mit denen der Learner trainiert werde, nachvollziehbar und dessen Code klar geschrieben seien, führe das Wechselspiel zwischen beiden zu nicht mehr nachvollziehbaren Komplexitäten. Dies sei bei der automatischen Spam-Erkennung beispielsweise der Fall. Diese operiere dadurch, dass jedes Wort in einer E-Mail daraufhin bemessen werde, inwieweit es mit Spam assoziiert sei. Kleine Unterschiede und Schlüsselwörter bilden – so Burrell – dann den Gradmesser dafür, ob eine E-Mail in den Spam-Ordner wandert oder nicht. Da bei diesem Filterungsprozess keine semiotische, geschweige denn eine semantische, also auf Bedeutungen abzielende Analyse, und schon gar nicht eine Narrationsanalyse zur Anwendung komme, die zum Beispiel Rückschlüsse auf die Intentionen des Verfassers der betreffenden E-Mails vornehmen würde, operiere die automatische Spam-Erkennung keineswegs für Menschen eingängig. Dies werde an der Liste derjenigen Wörter deutlich, die am stärksten mit einer Spam-E-Mail assoziiert seien: »our (0.500810), click (0.464474), remov (0.417698), guarante (0.384834), visit (0.369730), basenumb (0.345389), dollar (0.323674), price (0.268065), will (0.264766), most (0.261475), pleas (0.259571)« (Burrell 2016: 8). Unklar bleibe insbesondere die Gewichtung harmloser Wörter wie »visit« oder »want« als Indikatoren für Spam. Wenn also ein Machine Learning-Verfahren konsequent seine eigene Repräsentation der Klassifikationsentscheidungen aufbaue und Gewichtungen zwischen den

selbstgenerierten Klassifikationen so lange manipuliere, bis die Lerndaten die gewünschten Ergebnisse hervorbrächten, geschehe dies ohne Rücksicht auf die menschlichen Verstehensmöglichkeiten. Deshalb spricht sich Burrell generell dafür aus, Machine Learning nicht bei kritischen Anwendungen zu verwenden, und dort, wo Machine Learning zur Anwendung kommt, stets zu prüfen, ob dadurch unmittelbare oder mittelbare soziale Diskriminierungen ausgelöst werden.

Einen anderen Aspekt von Opakheit thematisieren Bechmann und Bowker (2019), wenn sie auf die klassifikatorische Arbeit rund um die neue KI abheben. Sie differenzieren hierbei drei Schichten dieser Arbeit: (1) Die Datensammlung selbst kann durch Datenklassifikation mitgeprägt sein; (2) bei der Datenreinigung findet offensichtlich eine Klassifikation in brauchbare und nicht-brauchbare Daten statt; sowie (3) das Trainieren des Learners kann ebenfalls als klassifikatorische Arbeit begriffen werden. Bei standardisierten Klassifikationssystemen würden so beispielsweise kulturelle Unterschiede mit zum Teil fatalen Folgen vernachlässigt – etwa, wenn Minderheitskulturen benachteiligt würden; so geschehen bei einer ML-Anwendung zur Bilderkennung, welche People of Color fälschlicherweise als Gorillas klassifiziert habe. Gerade durch die Hidden Layers bei Deep Learning werde es schier unmöglich, die vorgenommenen Kategorisierungen logisch aufzuschließen. Dieses Problem spitzt sich für die Autor:innen bei Unsupervised Machine Learning zu, da dort scheinbar induktive Klassifizierer zur Anwendung kommen. So bilde die hier verwendete Latent Dirichlet Allocation (LDA) ein Unsupervised Machine Learning-Verfahren, das insbesondere für semantische Analysen von Textdaten eingesetzt werde. Genau besehen, stelle sich dieses sogenannte unsupervised Modell als ausgeprägt supervised heraus, da angefangen von der Festlegung der Topics, über die Datenreinigung bis hin zu einem Vorverständnis für als bedeutungsvoll gehaltene Cluster ein hoher Grad menschlicher Kontrolle in das Verfahren Eingang gefunden habe. Durch derartige klassifikatorische Arbeit jedenfalls würden durch Machine Learning-Verfahren regelmäßig falsche Exkludierungen genauso wie falsche Inkludierungen stattfinden – mit den oben beispielhaft näher umrissenen diskriminierenden Konsequenzen für Minderheiten und andere durch Klassifikationen marginalisierte Gruppen.

Dieser Ausblick auf diese fünf zentralen Topoi ist natürlich keineswegs erschöpfend. Auch wäre ein systematisches Anknüpfen an all diese Punkte jenseits der Möglichkeiten eines Sammelbandes. Aus der White'schen Perspektive sind ohne Frage die Aspekte der Kontrolle und der Opakheit das Offensichtlichste. Der Themenbereich um das Verhältnis von Kontrolle und Machi-

ne Learning trägt sogar einen von Whites Grundbegriffen in der Bezeichnung, auch wenn Mackenzie aus einer anderen theoretischen Richtung argumentiert. Die Opakheit der KI-Systeme stellt ein anderes aus soziologischer Perspektive enorm relevantes Phänomen dar. Mit Blick auf die empirischen Gegenstände wird aber auch deutlich, dass der Realexperimentcharakter der Anwendungen nicht zu übersehen ist.

4. Die Beiträge des Bandes

Dementsprechend schließt der Sammelband an einige der vorher skizzierten Problembereiche explizit an und widmet sich anderen eher indirekt. Im ersten Text von Roger Häußling und Marco Schmitt wird für den Gegenstand eine relationale theoretische Perspektive im Anschluss an White entwickelt, einschließlich der Klärung zentraler Begriffe, ihres Zusammenhangs und ihre Funktion im Anwendungskontext der Künstlichen Intelligenz.

Im Anschluss daran fokussiert der Sammelband auf einzelne Themen, die Künstliche Intelligenz als gesellschaftliche Herausforderung betrachten. Zu diesem Zweck ziehen die jeweiligen Autor:innen eine Reihe von zwischen 2019 und 2021 geführten Expert:innen-Interviews mit einschlägigen Forscher:innen an der RWTH Aachen und darüber hinaus für ihre Texte zu rate.

Bei den behandelten speziellen Themen erfolgt der Anschluss insbesondere an die Topoi der Kontrolle und der Opakheit. Ein Text widmet sich der Frage nach den Vorgängen, die sich eher der Beobachtung entziehen und solchen die aktiv ins Licht gerückt werden. Marco Schmitt und Christoph Heckwolf diskutieren in ihrem Text »KI zwischen Blackbox und Transparenz« der wichtige Widerspruch zwischen »blackboxing« und »explainability« als unterschiedliche Kontrollprojekte. Sie unterscheiden im Anschluss an White zwischen Kontrolle, Kontrollversuchen und Kontrollprojekten und zeigen die Transparenzproblematik empirisch auf unterschiedlichen Ebenen (Daten, Trainingssetting, Personen und mathematische Grundlagen) auf. Hierbei greifen sie zur Interpretation auch auf die White'sche Systematik von Getting Aktion und Blocking Aktion zurück, um die Strategien der Forschenden und Entwickelnden zu fassen.

Darüber hinaus nimmt ein Text den Forschungs- und Arbeitsprozess mit diesen Algorithmen in seiner kulturellen Verfasstheit in den Blick. Wie wird hier gelernt bzw. wie wird das Lernen verteilt und wie werden die Vorgehens-

weisen gerahmt. Dieser Frage gehen Claudius Härpfer und Nadine Diefenbach in ihrem Text »Zur Kunst des Lernens« nach, indem sie in Anlehnung an Herbert Simon und White ein Modell künstlichen Lernens entwickeln. Dieses Modell wenden sie auf die beiden Lernformen, das Supervised Learning und Unsupervised Learning, an, um unter Rückgriff auf das Interviewmaterial zu zeigen, wie und unter welchen Voraussetzungen sich Lernstories entwickeln können. Wobei am Ende deutlich wird, wie viel menschliches Bauchgefühl und Hintergrundrauschen in den Lernvorgängen trotz aller Metrisierung vorhanden ist.

Dem folgt die Frage nach der Macht der Daten, also konkret die Frage nach der Kontrolle, die hier ausgeübt wird bei der Formierung der Daten ebenso wie beim Einsatz der Ergebnisse. Um dieser Frage techniksoziologisch gehaltvoll nachzugehen, greifen Philip Roth, Matthias Dorgeist und Astrid Schulz in ihrem Text »Deep Learning Techniken als Boundary Objects zwischen Entwicklungs- und Anwendungsfeld« auf den Begriff der Boundary Objects zurück, als die sie Deep Learning Technologien begreifen. Diese Grenzobjekte haben die Eigenschaft, in Entwicklungsvorgängen strukturierend zu wirken und so das Verhältnis von Entwickler:innen und Anwender:innen dieser Technologien zu beeinflussen. Um aus dieser Perspektive das Wechselspiel von Entwicklung und Anwendung anhand zweier Projekte aus den Lebenswissenschaften zu skizzieren, stützen sie sich auf Feld- und Praxistheoretische Konzepte, um diese im Nachklang am White'schen Ansatz zu messen.

Schließlich werden drei zentrale Anwendungskontexte mit erheblichen gesellschaftlichen Implikationen in den Blick genommen. Hierbei zeichnet sich deutlich insbesondere der Realexperimentcharakter der Machine Learning-Anwendungen ab. Da ist zum einen der Bereich von Medizin und Gesundheit, in dem man sich einerseits erhebliche positive Effekte vom Einsatz künstlicher Intelligenz erhofft, und andererseits gleichzeitig einem Bündel an ethischen und rechtlichen Bedenken gegenübersteht. Dieses Spannungsfeld betrachten Dhenya Schwarz und Tabea Bongert in ihrem Text »Über Identitäten und Selbstverständnis medizinischen Personals in Zeiten künstlicher Intelligenz und Algorithmierung« aus der Andwender:innenperspektive, indem sie das Selbstverständnis von Ärzt:innen und deren Verhältnis zu Patient:innen im sich wandelnden klinischen Arbeitsalltag in den Blick nehmen. Unter Rückgriff auf die Theorie Whites und Arbeiten Andrew Abbotts rekonstruieren sie die Institution Medizin im Spannungsfeld jener stabilisierenden und innovierenden Faktoren und geben mit Hilfe von einigen Interviews Einblicke in den Wandel der Ärzt:innenrolle.

Als zweiter Anwendungskontext dient die Perspektive der technischen Start-up-Kultur, die sich um neue datengetriebene Plattform-Konzepte dreht, ohne die Künstliche Intelligenz in ihrer aktuellen Form nicht möglich wäre und für die Künstliche Intelligenz dementsprechend ein verheißungsvoller Zukunftsmarkt ist. Tim Franke, Niklas Strüver und Sascha Zantis betrachten in ihrem Text »Plattform und jetzt?« die Gründungsstories digitaler Start-up-Unternehmen. Hierzu nutzen sie von theoretischer Seite Whites Story-Begriff und Gedankenfiguren des Solutionismus, der sich in jener schnelllebigen Startup Welt als Rechtfertigungsordnung einer Polis der Solution etabliert hat, die Start Up-Gründer:innen dazu bringt, sich den Unsicherheiten und Widrigkeiten einer Gründung auszusetzen. Als empirische Basis dienen acht Interviews, die im Rahmen des BMBF-geförderten Projekts INDIZ erhoben wurden.

Den Abschluss der Aufsätze machen Axel Zweck und Thomas Werner, die in ihrem Beitrag das Thema Künstliche Intelligenz aus Perspektive der Zukunftsforschung beleuchten. In einem Abriss der historischen Entwicklung künstlicher Intelligenz zeigen sie, wie diese ihre Rolle ändert. Vom Beginn als große Utopie hin zur schwachen KI, die Insellösungen bietet und mit der Zeit immer leistungsfähiger wird, um schließlich nicht mehr bloß reines Instrument zu sein, sondern zu einer Art Akteur:in im Forschungsprozess aufzusteigen und die Art zu Forschen verändert. Letzteres gilt insbesondere für die Zukunftsforschung, deren Prognoseverfahren mit den neu entstandenen Möglichkeiten einerseits ausgebaut werden können, andererseits stellt sich die Frage der Qualitätssicherung.

Schließlich greifen die Herausgeber den Faden nochmals auf und ziehen ein systematisierendes Resümee um anschließend – wie bereits erwähnt – den aktuellen Entwicklungen rund um ChatGPT Rechnung zu tragen und den Chatbot selbst in einem Interview zu Wort kommen zu lassen.

Dieser Sammelband ist im Arbeitsbereich des Lehrstuhls für Technik und Organisationssoziologie der RWTH Aachen University aus einem gemeinsamen techniksoziologischen Interesse an den Entwicklungen der neueren KI heraus entstanden. Die ersten, oben erwähnten Interviews wurden in diesem Zuge 2019 geführt. 2021 erschien das erste gemeinsame Working Paper, das versucht hat, die White'sche Perspektive auf ein Recommender-System anzuwenden (Häußling et al. 2021). Im Anschluss daran, nahm dieser Band Form an. Wir danken allen Beteiligten Autor:innen für ihre Diskussionsfreudigkeit und ihre Bereitschaft, tief in ein Thema abzutauchen, das mit dem Tagesgeschäft nicht immer einfach in Einklang zu bringen war. Darüber hinaus un-

terstützten uns bei der Durchführung der Interviews und in der konzeptionellen Phase des Bandes Annika Fohn, Julia Kolb und Jacqueline Lemm. Von Seiten des Verwaltungsteams seien Christa Siebes und Marlon Steffens genannt, die zu unterschiedlichen Zeitpunkten wertvolle Beiträge lieferten. Finja Bersch und Josias Bruderer schließlich gingen uns bei der Formatierung und Finalisierung für den Druck zur Hand.

5. Literatur

Bächle, Thomas C./Ernst, Christoph/Schröter, Jens/Thimm, Caja (2018): »Selbstlernende autonome Systeme? Medientechnologische und medientheoretische Bedingungen am Beispiel von Alphabets Differentiable Neural Computer (DNC)«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien. Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript, S. 167–192.

Bechmann, Anja/Bowker Geoffrey C. (2019): »Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media«, in: Big Data & Society, S. 1–11.

Bleher, Hannah/Braun, Matthias (2023): »Wissen und nicht wissen. ChatGPT & Co und die Reproduktion sozialer Anerkennung«, in: Forschung & Lehre 30, S. 260–261.

Burrell, Jenna (2016): »How the machine ›thinks‹: Understanding opacity in machine learning algorithms«, in: Big Data & Society 3, S. 1–12.

Engemann, Christoph (2018): »Rekursionen über Körper. Machine Learning-Trainingsdatensätze als Arbeit am Index«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien. Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript, S. 247–268.

Foucault, Michel, (1993): »Technologien des Selbst«, in: Luther H. Martin/Huck Gutman/Patrick H. Hutton (Hg.), Technologien des Selbst, Frankfurt a.M.: S. Fischer, S. 24–62.

Häußling, Roger/Franke, Tim/Härpfer, Claudius/Roth, Philip/Schmitt, Marco/Strüver, Niklas/Zantis, Sascha (2021): »Mendelian und das Erklärungspotential der Theorie von Identität und Kontrolle. Ein techniksoziologischer Blick auf Recommender-Systeme«. Working Paper des Lehrstuhls für Technik- und Organisationssoziologie 02, Aachen: Lehrstuhl für Technik- und Organisationssoziologie, Institut für Soziologie, RWTH Aachen. [https://publications.rwth-aachen.de/record/822003/files/822003.pdf]

Heiberger, Raphael (2022): »Applying Machine Learning in Sociology: how to Predict Gender and Reveal Research Preferences«, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 74, S. 383–406.

Mackenzie, Adrian (2017): Machine Learner. Archaeology of a Data Practice, Cambridge: MIT.

Mainzer, Klaus (2019): Künstliche Intelligenz. Wann übernehmen die Maschinen?, Berlin: Springer.

McCarthy, John/Minsky, Marvin/Rochester, Nathaniel/Shannon, Claude E. (2006): »A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955«, in: AI Magazine 27 (4), S. 12–14.

Mökander, Jakob/Schroeder, Ralph (2021): »AI and social theory«, in: AI & Society (2021), https://doi.org/10.1007/s00146-021-01222-z

Molina, Mario/Garip, Filiz (2019): »Machine Learning for Sociology«, in: Annual Review of Sociology 45, S. 27–45.

Nehlsen, Johannes/Fleck, Tilmann (2023): »Zulässige Hilfsmittel für Hochschulprüfungen? Rechtliche Aspekte von ChatGPT«, in: Forschung und Lehre 30, S. 262–264.

Parisi, Luciana (2018): »Das Lernen lernen oder die algorithmische Entdeckung von Informationen«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien. Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript, S. 93–113.

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

Peirce, Charles S. (1976): The New Elements of Mathematics: Vol. 4: Mathematical Philosophy, The Hague: Paris Mouton Publishers.

Rosengrün, Sebastian (2021): Künstliche Intelligenz. Zur Einführung, Hamburg: Junius.

Stilgoe, Jack (2018): »Machine Learning, social learning and the governance of self-driving cars«, in: Social Studies of Science 48, S. 25–56.

Sudmann, Andreas (2018): »Zur Einführung. Medien, Infrastrukturen und Technolohgien des maschinellen Lernens«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien. Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript, S. 9–23.

Zweig, Katharina (2019): Ein Algorithmus hat kein Taktgefühl. Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können, München: Heyne.