Aspekte einer Soziologie künstlicher Intelligenz

Es ist Anliegen dieses Kapitels, maschinelles Lernen als Gegenstand soziologischer Forschung zu erschließen. Der teils anekdotisch vermittelten ersten Annäherung an das Thema dieser Studie soll nun ein differenzierter Überblick zu verschiedenen Phänomenaspekten von ML erarbeitet werden. Das Kapitel ist in vier Abschnitte untergliedert. Diesen folgt ein zusammenfassendes Zwischenfazit.

Der erste Abschnitt erörtert, in welchen Hinsichten ML als eine sozialtheoretische Herausforderung zu verstehen ist. Ausgehend von einer Einordnung aktueller sozialtheoretischer Bezugnahmen auf ML wird mit Bezug auf die Aspekte Wahrnehmung, Kommunikation und Wissen exemplarisch dargestellt, welche Lücken sozialtheoretischen Denkens die Gegenwart von ML offenbart.

Im Anschluss daran wird der für soziologische Diskussionen um ML prägende Begriff des Algorithmus in den Blick genommen. Dabei wird ML von früheren Formen sog. KI unterschieden und darüber vermittelt näher bestimmt. In Annäherung an die rechentechnische Operationslogik von mit ML assoziierten Verfahren wird ebenso erörtert, worauf sich das *Lernen* in *maschinellem Lernen* gemeinhin bezieht. Dabei werde ich unter Rückgriff auf Warren Weavers Überlegungen zum Begriff der Information die These entwickeln, dass der Einsatz von ML als (im rechentechnischen Sinne der Informatik) informationsverarbeitende und aus der Perspektive von Beobachter:innen zugleich ›sinnverarbeitende‹ Maschine für ebendiese Beobachter:innen mit der Emergenz eine unüberwindbaren epistemische Lücken einhergeht, die das ›Verstehen‹ von Modelloutputs verkompliziert, wenn nicht verunmöglicht.

Der anschließende Abschnitt erörtert mit Blick auf bisherige soziologisch relevante Forschung zu ML die Spannung zwischen probabilistischen und kausaltechnischen Eigenschaften, die ML als empirischen Gegenstand auszeichnen. Es wird dabei ein argumentativer Zusammenhang aufgebaut zwischen der für ein angemessenes (sozial-)theoretisches Verständnis von ML zentralen stochastischen Modellierung von datenförmig repräsentierten Zusammenhängen und den ambivalenten Beurteilungen, die praktische Einsätze von ML durch Beobachter:innen erfahren. Weil ML gleichermaßen Eigenschaften von Kausaltechnik und Kommunikation in sich vereint, zeitigt sein praktischer Einsatz eine Vielzahl von teils wider-

sprüchlichen sozialen Effekten. Zur Begründung dieser Ansicht wird die probabilistische Logik stochastischer Errechnungen von Verteilungen als Heuristik für das Verständnis algorithmischer Modelle herangezogen. Diese Logik wird sich als zentral für ein soziologisches Verständnis von ML erweisen, da sie die Genese jener Modelloutputs, die schließlich von Beobachter:innen als Analysen und Vorhersage über die datenförmig repräsentierten Zusammenhänge beobachtet werden, zu beschreiben vermag.

Danach verschiebt sich der Blick von den internen Operationen der Maschinen auf die externen Bedingungen ihrer Konstitution. Deren Erörterung nimmt die räumlichen wie zeitlichen Extensionen von ML in den Fokus. Es zeigt sich dabei, dass die Genese von ML-Anwendungen als Akkumulationsvorgang, der über verschiedene Orte hinweg in verschiedenen zeitlichen Phasen unterschiedliche Ressourcen mobilisiert, deren Synthese erst ermöglicht.

Im abschließenden Zwischenfazit wird ML dann schließlich zusammenfassend vor allem hinsichtlich der Schwierigkeiten und Unwägbarkeiten gekennzeichnet, die Versuche der Bestimmung des Sinns von Modelloutputs mit sich bringen. ML-vermittelte Sozialität, so die für die weitere Untersuchung leitende These, erzeugt in praktischen Anwendungszusammenhängen eine charakteristische epistemische Lücke, die das soziale Geschehen mit sowie rund um ML grundlegend prägt.

Eine ambivalente Verfassung

Die Bandbreite soziologischer Beurteilungen von ML ist offenkundig vielfältig und keinesfalls von Konsens geprägt. Dissens findet sich nicht nur hinsichtlich normativer Bewertungen, auch wenn diesbezügliche Kontroversen die Diskussion insbesondere um die Zukunft von ML vor allem in (post-)massenmedialen Zusammenhängen den Ton anzugeben scheinen. Auch mit Blick auf (im weiteren Sinne[1]) soziologische Theoriebildung zu ML finden sich durchaus grundlegend verschiedene Perspektivierungen von ML als sozialem Phänomen. Angesichts des Anliegens dieser Untersuchung erscheint eine Auseinandersetzung mit den entsprechenden theoretischen (geistes- und sozialwissenschaftlichen wie auch philosophischen)

1 Mit dieser Formulierung will ich zum Ausdruck bringen, dass es mir primär um soziologische Theoriebildung im Sinne des Beschreibungsgehalts und nicht in erster Linie um explizite disziplinäre Zuordnung geht, weil soziologisch instruktive Beiträge zu ML sich meiner Erfahrung nach vor allem auch in Publikationen finden, die formal nicht unbedingt dem soziologischen Fachdiskurs zugerechnet werden. Entsprechend werde ich in meinem Argumenten im Verlauf der Untersuchung neben soziologischer auch solche Literatur berücksichtigen und auswerten, die ›offiziell‹ etwa den Diskursen der Medienwissenschaft, der Kulturwissenschaft, den STS oder der Technik- und Medienphilosophie zugehörig ist.

Auseinandersetzungen offensichtlich aussichtsreicher als eine Vertiefung der normativ orientierten Debatten um ML.

Der Theoriediskurs zu ML umfasst einerseits Positionen, die ML entweder explizit theoretische Relevanz absprechen (»fancy way of calculating an average«, Galloway 2021) oder gleiches implizit vollziehen, indem sie ML keiner gesonderten Betrachtung unterziehen und gewissermaßen begrifflich einebnen in einen größeren Phänomenzusammenhang wie etwa ›Digitalisierung‹/›Digitalität‹ oder ›Big Data‹/›Datafication‹ (van Dijck 2014)[2], der ML dann gleichsam unter der Hand miterkläre. Die anhaltende dramatische Inszenierung von Kontroversen um ML-Utopien und (vor allem) -Dystopien im Modus der Spekulation stellt aus der Perspektive solcher Positionen vermutlich ein interessantes Spektakel, mit Blick auf seinen theoretischen Gehalt im Grunde aber wohl einen »total scam« (ebd.) dar, als dessen Funktion vor allem die Invisibilisierung der materialen Effekte des Einsatzes von ML zu verstehen sei.[3]

Am anderen Ende des diskursiven Spektrum erscheinen dementgegen Positionen, die für Verständnisversuche von ML eine doppelte Differenz geltend machen. Einerseits findet sich eine Distanzierung zur Analogisierung von ML und Mensch. Folglich wird Abstand genommen von theoretischen Beschreibungen, die allzu sehr an typische Charakterisierungen menschlicher Fähigkeiten, Tätigkeiten und anderen eng mit Menschlichkeit verbundenen Konzepten, da das »alien subject of AI« (Parisi 2019) einen post-anthropozentrischen sozialtheoretischen Zugriff nötig mache. Andererseits geht es um die Differenz zu bekannten Formen digitaler Technik. ML zeichne sich im Gegensatz zu im Grunde allen vorangegangenen technischen Entwicklungen gerade durch einen eigentümlichen Sinnbezug aus, der eng mit der rechnerischen Modellierung von datenförmig abgebildeten Sinnzusammenhängen zusammenhänge. Andererseits finden sich in diesem Zusammenhang theoretische Positionen, die das Neue in ML bzw. die genuine Eigensinnigkeit von sich in ML-Verfahren manifestierenden Formen von »computational reason« (Cavia 2022) oder von im Zuge von deren praktischer Anwendung emergenter »algorithmischer Sozialität« (Seyfert 2023) betonen. Solche Begriffe fungieren als

2 Oder aber in gesellschaftstheoretischer Absicht ›Technokapitalismus‹/'Plattformkapitalismus‹ (vgl. Srnicek 2018, Zuboff 2018) als spezifische Varianten von ›digitaler Gesellschaft‹ (Nassehi 2019). Auf dieser Seite (des Diskurses) erscheint ML mithin primär als bloße Fortschreibung von hinlänglich Bekanntem unter leicht veränderten Vorzeichen, etwa dem Industriekapitalismus samt Fließbandprinzip (vgl. Pasquinelli & Joler 2021). Im Sinne eines Fokussierung auf das Kernanliegen der vorliegenden Untersuchung werde ich von expliziten Auseinandersetzungen mit gesellschaftstheoretischen oder zeitdiagnostischen Thematisierungen von ML weitestgehend absehen.

3 Vgl. weiterführend zur Funktion von Projektionen utopischer wie dystopischer Zukunftsszenarien im Kontext von KI/ML Burrell 2023 sowie im Zusammenhang mit digitalen Technologien allgemeiner Dourish & Bell 2014.

Differenzmarkierungen – in diesem Fall offensichtlich zu üblichen Verständnisse von »reason« und »Sozialität«[4] – und mithin als Hinweise auf Leerstellen der bisherigen theoretischen Theoriearbeit. In sozialtheoretischer Thematisierung dieses Umstandes finden sich Ansätze, die explizit einen Bedarf an begrifflicher Arbeit artikulieren und sich vor allem in Kritik des ›Theoriebestands‹ üben, indem sie die Eigenlogik rechentechnischer Phänomene betonen (vgl. Bratton & Agüera y Arcas 2022). Auf eine griffige Formel gebracht: »*computation is computation*« (Fazi 2018, S. 187).[5] Beiden letzteren Positionen ist die Ansicht gemein, dass ML eine genuine Herausforderung für die Theoriebildung darstellt und nicht etwa nur alte Probleme in neuen Gewändern vorführt.

Unter ähnlichen Vorzeichen lassen sich Luciana Parisis Ausführungen zum »alien subject of AI« (2019) nachvollziehbar machen. Laut Parisi geht die gegenwärtige Verbreitung von ML mit einer Automatisierung kognitiver Tätigkeiten einher, was mithin der Ausbreitung eines »alien space of reasoning« (S. 30) gleichkomme. Dieser sei geprägt durch den für Beobachter:innen befremdliche Charakter dessen, was sie »machine thinking« (S. 31) nennt. Die Verbreitung von stochastischen Modellierungen zur automatischen Generierungen von sinnförmig verstehbaren Outputs für kommunikative Anlässe verheiße nicht weniger als eine Krise »crisis of conscious cognition« (S. 28). Es verbreite sich innerhalb der vertraut geglaubten sozialen Wirklichkeit zudem ein für menschliche Beobachtungen opaker »space of machine communication« (S. 31). Die aus diesen Entwicklungen resultierende Krise menschliche Subjektivität hat laut Parisi zwei wesentliche Bezugspunkte: Handeln und Wahrnehmung. Im Hinblick auf beide erweist sich »alienness« als zentrales Charakteristikum. Ungewöhnliche Outputs von ML-Modellen (vgl. etwa Bucher 2017) können Erwartungen ihrer Beobachter:innen destabilisieren und erschweren gleichermaßen Verstehen sowie adäquate Reaktionen, weil das Befremdliche an ihnen kaum rationalisierend durch Erklärungen eingehegt, d.h. verstehbar gemacht werden kann.

Beobachter:innen von ML-Anwendungen könnten im Versuch, deren Outputs als Kommunikation zu verstehen, nachgerade einen »epistemic shock« (Roberge &

4 Die entsprechend ›non-computational‹ bzw. ›nicht-algorithmisch‹ verfasst sein müssten, was angesichts der langen Kulturgeschichte des Rechnens sowie spezifisch der Algorithmen (vgl. etwa Pasquinelli 2019) eine anspruchsvolle Aufgabe darstellt.

5 Das Insistieren darauf, dass »computation« ein eigenlogisches Phänomen darstellt, impliziert zugleich Skepsis gegenüber gängigen Metaphern und Analogien: »computation is neither nature nor artifice« (ebd.) und gleichsam »neither about mathematical idealism nor physical empiricism« (Fazi 2016). Kontextabhängig wird »computational« in der vorliegenden Studie meist als »rechentechnisch« übersetzt, weil »computation« hier primär in ihrer Manifestation als Technik Gegenstand ist. Darüber hinaus ist vor allem relevant, dass jede »computation« zwangsläufig eine bestimmte Rechnung darstellt, was sich in Form typischer Vorsilben sprachlich abgebildet: *Er*rechnung, *Be*rechnung.

Castelle 2021, S. 2) erfahren. Dies sei sozial- wie gesellschaftstheoretisch keine Bagatelle, schließlich seien die jüngeren Entwicklungen um ML als dessen überaus folgenreicher »claim to meaning itself« (ebd., S. 7) zu verstehen. In dessen Folge verbreite sich eine spezifische, ML eigene, Sozialität, deren Erforschung die wesentliche Aufgabe für eine ›Soziologie maschinellen Lernens‹ oder »Critical AI Studies« (Raley/Rhee 2023) im weiteren Sinne darstellt. Es müsste dann in mancher Hinsicht neu angesetzt werden, etwa wenn in ML-Anwendungen deren empirische Auswirkungen weit von den Absichten und Erwartungen, der verantwortlichen Designer:innen, Ingeneur:innen, Programmierer:innen usw. abweichen, was gut dokumentiert ist (vgl. etwa Broussard 2018). An einer solchen Stelle wäre der Hinweis darauf, dass dies techniksoziologisch betrachtet keinesfalls außergewöhnlich sei (vgl. etwa die Beiträge in Bijker/Law [Hg.] 1992, bes. Akrich 1992) zwar zutreffend, aber als Erklärung nicht ausreichend, da der konkrete Charakter des Auseinandertretens von Erwartung und praktischem Resultat im Falle von ML-Anwendungen vielleicht ein genuin eigener, bislang nicht bekannter sein könnte. Der unbestreitbaren Prägnanz der empirischen Effekte von ML – die womöglich bis zu den Grundbedingungen von Sozialität reichen (vgl. Sujon und Dyer 2020) –, stehen folglich konzeptuelle Bemühungen gegenüber, die mitunter bis zur Infragestellung sozialtheoretischer Grundbegriffen wie Kognition, Interaktion, Handlung (bzw. *agency*[6]) oder kollektivem Verhalten geführt haben (vgl. etwa Yolgörmez 2021, Hayles 2017, Borch 2022).

Die bisherige sozialtheoretische Auseinandersetzung mit ML hat mangels Alternativen[7] gewissermaßen vor allem einen indirekten und – differenzlogisch formuliert – negativen Charakter. Indirekt wären etwa Vorgehen zu nennen, die zwar zu zeigen imstande sind, *dass* ML eine Herausforderung für die sozialtheoretische Begriffsarbeit bedeutet. Doch mit dem Hinweis darauf, welche Begriffe angesichts von ML überdacht werden sollten, ist allerdings noch keine positive Aussage über ML getroffen. Als negativ lassen sich in diesem Sinne Vorschläge der Charakterisierung von ML kennzeichnen, die vor allem zeigen, wie es *nicht* zu verstehen sei.

Dabei sind in den letzten Jahren instruktive begriffliche Ansätze entstanden, die sich die Spannung der disparaten Eigenschaften von ML – oben als ›doppelte Differenz‹ beschrieben – explizit zu eigen machen. Bei Formeln wie »non-conscious

6 Die vorliegende Arbeit verzichtet auf eine eindeutige Übersetzung von *agency*, da die Polysemie des Konzepts eine solche ohne Verzerrungseffekte nicht erlaubt. Wo geboten, wird versucht, kontextabhängig begriffliche Erläuterung zu leisten. *Agency* bezieht sich auf Wirkungsmacht in einem Spektrum, dass in unbestimmter Form menschliche wie nicht-menschliche Vermögen umfasst und deshalb weder allein etwa als ›Handlungsträgerschaft‹ (zu ›aktivistisch‹) noch ›Aktivitätspotenzial‹ (zu ›schwach‹) adäquat erfasst ist. Für eine ausführlichere Reflexion des Übersetzungsproblems vgl. Roßler 2007.

7 Dies im Sinne von Ogburns oben vorgestellter These eines »cultural lag«, vgl. oben Fn. 31.

cognition« (Hayles 2017), »algorithmic thought« (Fazi 2021) oder »artificial communication« (Esposito 2017, 2022, siehe aber auch bereits Esposito 1993) handelt es sich um widersprüchlich anmutende Komposita, die zugleich Nähe und Distanz zu sozialtheoretisch Vertrautem markieren. Durch eine ›humanistische Brille‹ betrachtet erscheinen alle drei Ausdrücke als Oxymora, die enge Assoziationen mit Menschlichkeit (conscious cognition – thought – communication) durch gegenteilig anmutende Attribute (non-conscicous – algorithmic – artificial) konterkarieren. Es handelt sich um Formeln, die mit dem Bezug auf das Menschliche zugleich das semantische Register um das notorische *moving target* KI/ML aufrufen, in dessen historischem Wandel ebendieser Bezug wiederholt zum Maßstab für die Entwicklung sogenannter Künstlicher Intelligenz erklärt wurde.

Die rhetorische Figur des Oxymorons erlaubt die Repräsentation einer grundlegend ambivalenten Verfassung, wie sie von den drei Theoretikerinnen mit unterschiedlichen Problembezügen beobachtet wurde: ML sei zu kognitiven bzw. ›denkähnlichen‹ Leistungen fähig, wie sie sonst nur von Menschen bekannt seien (Hayles, Fazi) und könne sich an Kommunikation beteiligen, wie es sonst nur Menschen könnten (Esposito). In allen drei Fällen markiert das zusätzliche Attribut indes zugleich einen Bruch mit ebendieser Suggestion, dessen theoretische Relevanz womöglich selbst die jener Suggestion übersteigen könnte, auf die er sich bezieht. Dies gilt insofern, als dass die gleichzeitige Identifikation und Desidentifikation in ML gewissermaßen in ein sozialtheoretisches Niemandsland führt, da es sowohl zu Menschlichen als auch zu anderen nicht-menschlichen Phänomenen auf Distanz gebracht wird.

Es ist diese eigentümliche konzeptuelle Disposition, die mein Interesse an ML als Gegenstand sozialtheoretischer Theoriebildung geweckt hat. Diese will ich mir zum Ausgangspunkt der vorliegenden Untersuchung machen, für die folglich die maßgebliche Forschungsfrage eine sehr grundlegende ist: Wie lässt sich ML adäquat soziologisch adäquat beschreiben? Insbesondere erhoffe ich mir dabei eine Form der theoretischen Beschreibung, die – jenseits der unhintergehbaren Referenz auf (nicht nur, aber mindestens: auch) menschliche Sozialität im Kontext von Sozialtheorie – die für ein soziologisches Verständnis von ML nicht auf den Rückgriff zur Mensch-Analogie angewiesen ist. Stattdessen verfolge ich das Anliegen einer positiven Bestimmung der Stellung (und mithin der Funktion) von ML in einer sozialen Wirklichkeit, in der sich – dies gewissermaßen als ultrakondensierte Variante der zentralen Einsicht soziologischer Forschung überhaupt – ohnehin nicht alles um *den* Menschen dreht. Ein derartiges Ziel zu verfolgen, scheint mir vielversprechend, weil damit einerseits womöglich neue Perspektiven die bekannten Kontroversen um KI/ML auf produktive Weise irritieren können. Andererseits scheint sich mir ein solches Projekt darüber hinaus auch dem Anliegen zu verpflichten, zur Behebung des »cultural gap« (im Sinne einer epistemischen bzw. theoretischen Krise) beizutragen, von dem es mithin qua eigener Prämissen selbst ausgeht.

In diesem Sinne nehme ich an, dass die weitereichenden Effekte von ML über verschiedene soziale Zusammenhänge hinweg wesentlich mit der epistemischen Herausforderung zu tun haben, mit der sich nicht nur der theoretische, sondern auch der praktische Umgang mit Outputs von ML-Anwendungen konfrontiert sieht. Diese sind, wie eingangs erläutert, grundlegend als statistische Datenanalysen (Dourish 2016) zu verstehen, auch wenn sie die Form von Texten oder Bildern annehmen. Die Gestalt eines konkreten Outputs durch das zuvor algorithmisch errechnete Modell eines Datensatzes lässt sich nicht durch eine genetische Kausalerklärung nachvollziehbar machen und kann angesichts der zugrundeliegenden rechentechnischen Vorgänge nicht als sinnhafter Prozess (etwa analog zu Wahrnehmung, Denken oder Kommunikation) vorgestellt werden.

Wie Louise Amoore ausführt, vermögen ML-Anwendungen ihrerseits indes subtile Veränderungen in der menschlichen Weltwahrnehmung zu bewirken, die ohne ML womöglich gänzlich unregistriert verblieben (vgl. Amoore 2020, S. 40f.). Derartige Datenverarbeitungsformen verschieben nachgerade die Schwellen des für menschliche Beobachtung Zugänglichen, wenn nämlich dabei produzierte Daten als Weltdaten Informationen über die Wirklichkeit liefern bzw. in die Kommunikation einführen, die es ohne ML schlichtweg nicht gäbe.[8] ML als *Computer Vision* könne zwar nicht selbst im eigentlichen Sinne ›sehen‹, doch spiele es eine durchaus relevante Rolle in der Regulierung der Sichtbarkeit sozialer Wirklichkeit. Menschliche Weltwahrnehmung hänge entsprechend zunehmend davon ab, was durch ML-basierte Sichtbarkeitsregimes der menschlichen Beobachtung verfügbar gemacht wird. Diese Überlegung bezieht sich weniger auf den Verdacht, dass mit *Computer Vision* in erster Linie ein Verschleierungsinstrument am Werk sei. Vielmehr zielt sie grundlegender auf die Feststellung, dass technisch vermitteltes Sehen (wie auch ›Einsicht‹ im weiteren Sinne) das Ergebnis kontingenter Prozesse ist.[9] Verschleierungen der Realität geht das grundlegende Problem der Wirklichkeitsvermittlung überhaupt voraus. Befremdliche Erfahrungen mit ML-Outputs

8 Dies gilt nicht exklusiv für ML, sondern für zahlreiche Formen von Digitaltechnik, die mit Armin Nassehi gerade dort anfange, »wo sich die Welt in Daten repräsentieren lässt, um Muster und Strukturen zu erkennen, die mit bloßem Auge und den Wahrnehmungs- und Rechenkapazitäten des natürlichen Bewusstseins nicht erfasst werden können« (2019, S. 230). Noch grundlegender ist das von Amoore aufgerufene Thema medientechnischer Effekte auf menschliche Wahrnehmung nicht einmal allein auf Digitaltechnik beschränkt, sondern verweist auf einen etablierten medien- und kultursoziologischen Forschungskomplex im Zusammenhang mit Fragen nach technischer (Re-)Produzierbarkeit und Vervielfältigung (vgl. Schrage & Hieber 2007).

9 Der amerikanische Mark B. N. Hansen nimmt ähnliche Beobachtungen zum Ausgangspunkt seines Arguments, dass angesichts wachsender »environmental sensibility« (Hansen 2015, S. 8) durch technisch vermittelnde Arrangements diese im Rahmen phänomenologischer Medientheorie eine zentraleren Status gegenüber ›reiner‹ menschlicher Wahrnehmung erhalten sollten. Im Rückgriff auf Alfred North Whitehead reflektiert Hansen in seiner Studie

zeugen in diesem Sinne keineswegs zwangsläufig von Manipulationsversuchen oder technischen Störungen, sondern lassen sich gleichsam schlicht als Ausdrücke der epistemischen Schockwirkung von ML plausibilisieren (vgl. Roberge & Castelle 2021).

Diese Disposition – unsichtbare Veränderungen im Register der Wahrnehmung gepaart mit Fremdheitserfahrung im sozialen Kontakt – legt eine Vielzahl von Möglichkeiten nahe, die ML in konkreter sozialer Gestalt *potenziell* realisieren kann. Man könnte sich hier an Charakterisierungen von digitalen Computern als ›general-purpose machines‹ erinnert fühlen, die auf Charles Babbages »Analytical Engine« von 1837 sowie Ada Lovelaces Nachweis von deren algorithmischer Programmierbarkeit zurückgehen (vgl. Misa 2015 sowie Fazi 2018, Kap. 1). Auf dieser Argumentationsebene tritt vor allem das prinzipielle Potenzial von ML für eine Vielzahl von Anwendungen (oder »purposes«) als Kennzeichen der beschriebenen Disposition in den Blick.

Dass ML hier vor allem hinsichtlich seiner Potenzialität beschrieben wird, hat wiederum mit dem negativen Charakter der Betrachtungsperspektive zu tun. Wie oben für den Theoriediskurs um ML bereits als typisch markiert, wird ML bzw. die menschliche Begegnung mit ML sozial als Differenz bestimmt: ein »alien subject« (Parisi), das für *veränderte* Wahrnehmung sorge (Amoore). Der folgende Abschnitt wird einen ersten Versuch der Überwindung dieses theoretischen Dilemmas darstellen, indem ich mich über eine Erörterung der Rolle von Algorithmen der technischen Operationslogik von annähern werde.

Über Algorithmen

Algorithmen werden häufig als Kernelement von ML ausgemacht, wenn nicht sogar als Untersuchungsgegenstand direkt mit ML oder KI identifiziert, sodass Forschung im Feld der ›Soziologie der Algorithmen‹ kaum klar von solcher unterschieden werden kann, die sich als ›Soziologie künstlicher Intelligenz‹ versteht. Die semantische Nähe könnte nahelegen, dass Algorithmen als eine Art analytischer Schlüssel für ein adäquates soziologische Verständnis von ML aufzufassen seien. Dieser Überlegung möchte ich im Folgenden nachgehen.

Da – wie bereits gezeigt wurde – ML als rechentechnisches Verfahren zentral auf der algorithmischen Modellierung von Zusammenhängen auf Datenbasis beruht, drängt sich die These auf, dass die operative Logik von ML als sozialem Phänomen im Grunde der Logik der Rechenoperationen von Algorithmen entsprechen müsste. Die Verbreitung von teils auch über Fachdiskurse hinaus popularisierten

---

den Umstand, dass »twenty-first century media« menschliche Erfahrungen beeinflussen, ohne dabei jedoch als solche überhaupt wahrnehmbar zu sein (vgl. ebd., S. 4).

Konzepten wie ›algorithmische Diskriminierung‹ (etwa Houben & Prietl 2023), ›algorithmische Herrschaft‹ (»algocracy«, Danaher 2016) bzw. »algorithmic regimes« (Jarke et al. 2024), »algorithmische Rationalität« (Mersch 2019) oder »algorithmische Sozialität« (Seyfert 2023) scheint indes ein weiteres Indiz dafür zu liefern, dass soziologisch jedenfalls ein enger Zusammenhang zwischen ML und Algorithmen anzunehmen ist. Formeln wie die genannten schreiben Algorithmen mitunter als adressierbaren sozialen Entitäten bestimmte Potenziale, wenn nicht gar Fähigkeiten zu (zu diskriminieren, zu herrschen). Zumindest halten sie den Einzug einer als algorithmisch bezeichneten Operationslogik in die soziale Wirklichkeit fest, deren Charakter sie dadurch verändert.

»Algocracy« und verwandte Losungen verweisen weniger auf Theoriedebatten als vielmehr auf eine veränderte empirische Lage, insoweit sie sich Forschung zu den beträchtlichen Auswirkungen des Einsatzes von Algorithmen etwa in der Finanz-, Sicherheits- oder Pharmaindustrie verdanken. Es lässt sich dabei nicht immer auf den ersten Blick beurteilen, inwiefern »Algorithmen« per se gleichbedeutend mit ML verwendet werden, auch wenn die Prävalenz beider Begriffe im soziologischen und verwandten Diskursen im Zeitraum seit etwa 2012 gleichermaßen signifikant zugenommen hat. Da Algorithmen technologischer Kern jedes Computerprogramms sind, fungieren sie häufig als Chiffre für Digitaltechnik überhaupt (vgl. Kitchin 2017, S. 15). Es scheint mir deshalb angebracht, den Stellenwert von Algorithmen in ML explizit zu erörtern, um im Kontext dieser Studie ein plausibel trennscharfes Verständnis von ML (gegenüber« Algorithmen« oder »Digitaltechnik«) zu etablieren. Dies ist nicht nur für die Operationalisierung von ML im empirischen Teil meiner Untersuchung unerlässlich, sondern vermag ebenso begriffliche Orientierung im mitunter unübersichtlichen diskursiven Geschehen zu stiften. Zuletzt ist darüber hinaus unbedingt zu prüfen, inwiefern Algorithmen bzw. ›das Algorithmische‹ Relevanz für das Anliegen dieser Studie haben.

Diskursiv werden beide Begriffe nicht selten austauschbar genutzt, wenn z.B. in zeitdiagnostischer Absicht von einer in Entstehung befindlichen »Society of Algorithms« (Burrell & Fourcade 2021) oder »Algorithmuskulturen« (Roberge & Seyfert 2017) die Rede ist und sich zeigt, dass beide semantische Schöpfungen empirisch wesentlich auf Zusammenhänge verweisen, die im Kontext dieser Studie als ML zu verstehen wären. Unabhängig von der Beurteilung solcher zeitdiagnostischer Formeln scheint im Fachdiskurs unstrittig, dass Algorithmen für die Soziologie gegenwärtig überaus relevante Untersuchungsgegenstände darstellen. Ihre soziologische Bestimmung hat allerdings ihre Tücken, etwa wenn Algorithmen als »performative in nature and embedded in wider socio-technical assemblages« (Kitchin 2017, S. 16) verstanden werden, was ihre Erforschung vor methodologische und methodische Herausforderungen stellt, auch weil in ihrem Einsatz mit nicht-intendierten Nebeneffekten zu rechnen ist, wobei im Sinne der Definition Rob Kitchins davon auszugehen ist, dass in praktischen Anwendungen von Algorithmen ein ganzes Netz

sozialer Beziehungen mobilisiert wird (vgl. ebd., S. 19).[10] Eine solche Konzeptualisierung führt zugleich vor, dass eine nicht hinreichend spezifizierte Bestimmung fraglich macht, was genau gemeint sein könnte – und was nicht –, wenn von Algorithmen die Rede ist, und ob die zugeschriebenen Attribute (»performative […] and embedded«) tatsächlich Algorithmen oder nicht etwa vielmehr das soziale Geschehen betreffen, aus dem sie hervorgehen, das sie begleiten oder für das sie sorgen.

In der vorliegenden Argumentation sind Algorithmen bislang nur randständig vorgekommen, etwa zur nicht weiter explizierten Charakterisierung rechnerischer Vorgänge wie etwa *algorithmischen* Modellierungen. Der folgende Abschnitt widmet sich ausführlicher dem Stellenwert von Algorithmen innerhalb der sich hier entwickelnden sozialtheoretischen Auseinandersetzung mit ML. Die bereits eingangs knapp eingeführte Unterscheidung zwischen ML und der auch als *Good Old Fashioned Artificial Intelligence (GOFAI)* bekannten *symbolischen KI* vermag sich hier als instruktiv erweisen. Dies gilt zum einen, weil die Rolle von Algorithmen in beiden Ansätzen fundamental unterschiedlich gelagert ist und zum anderen, weil ein über diese Einsicht vermitteltes Verständnis der mitunter trivialen Funktionen von Algorithmen deren Rolle entmystifizieren kann.

Diesem Vergleich soll eine Erörterung zum Begriff des Algorithmus vorausgehen, weil dies angesichts der mit dem gegenwärtig inflationärem kommunikativen Gebrauch einhergehenden Vervielfältigung möglicher Wortbedeutungen allemal geboten scheint. Es ist dies übrigens eine diskursive Lage, die vor zwanzig Jahren noch kaum abzusehen war, wurde von Algorithmen damals noch weitgehend exklusiv in Kreisen der Mathematik und Informatik gesprochen. Vom vor allem sozial- und geisteswissenschaftlich geprägten KI/ML-Diskurs einmal abgesehen, in dem – wie bereits angedeutet – allerdings nicht unbedingt immer klar ist, in welchem Sinne Algorithmen zu verstehen seien, hält dieser Zustand allerdings gewissermaßen an. Aus der ethnografischen Forschung ist bekannt, dass selbst die vermeintlichen Produzent:innen von Algorithmen nicht glauben, mit diesen operativ überhaupt zu tun zu haben, geschweige denn für deren Herstellung verantwortlich zu sein. Folglich werden Ethnograf:innen im Feld vage an andere Abteilungen des eigenen Betriebs weiterverwiesen, in denen womöglich tatsächlich an oder mit Algorithmen gearbeitet werde (vgl. Seaver 2017, S. 3).

Man könnte fragen, was es mit den Algorithmen selbst zu tun haben könnte, dass sie überall und nirgends zu sein scheinen. Schlichtweg überall, so ließe sich argumentieren, würden sie auftreten, wenn man ihnen mithilfe einer allgemeinen gefassten Definition auf die Spur kommen möchte. Klassisch verstanden als »well-

10 Kitchin identifiziert drei Herausforderungen für die Erforschung von Algorithmen: 1. Zugang zu ihrer Genese (Formulierung) zu gewinnen, 2. ihren heterogener Charakter und ihre soziale Einbettung angemessen zu erfassen, 3. der kontextuellen und kontingenten Entfaltung ihres Wirkens zu folgen (S. 20ff.).

defined computational procedure[s]« (Burke 2019, S. 1) muten Algorithmen auf den ersten Eindruck nicht sonderlich spektakulär an und ihr Dasein ›unter der Haube‹ von Digitaltechnik sorgt zumeist für wenig Aufsehen.[11] Jede Software besteht bekanntlich aus Sets von Rechenanweisungen; ein Umstand, der die gegenwärtige Faszination für Algorithmen kaum zu erklären vermag. Ist heute von »Algorithmuskulturen« (Roberge & Seyfert 2017) oder »algorithmischer Sozialität« (Seyfert 2023) die Rede, muss deshalb davon ausgegangen werden, dass sich mit solchen Begriffsschöpfungen nicht bloß wiederholt, was schon allgemein Prozessen der »Digitalisierung« oder der »Computerisierung« attestiert wurde. »Algorithmic Culture« (Galloway 2006) ist konzeptuell älteren Datums als der gegenwärtige ›KI-Sommer‹ und bezeichnete dies- und jenseits von KI schon vor 20 Jahren, wie mittels Computertechnik Menschen, Orte, Objekte und Ideen sortiert, klassifiziert und hierarchisiert werden (vgl. Striphas 2015). Von einer Rechenanweisung und selbst der Verkettung einer Vielzahl miteinander verschränkter Befehle kann wohl kaum eine gerade Linie gezogen werden zu den (wie prätentiös auch immer gearteten) Sorgen um dystopische Szenarien, wie sie in den erwähnten Imaginationen der nahenden Schwelle zur »Singularität« bzw. zu *Artificial General Intelligence (AGI)* artikuliert werden.

Allerdings lässt sich erkennen, dass es durchaus für ML zentrale Algorithmen gibt; jene nämlich, die als *lernfähig* verstanden werden (vgl. Parisi 2018).[12] Es handelt sich dabei um eine Eigenschaft, die in aller Regel auf Kosten der Transparenz der rechnerischen Vorgänge geht. Die resultierende Opazität der beteiligten Maschinen ist dabei genuin den Rechenvorgängen selbst zuzuschreiben und nicht etwa ›technischem Analphabetentum‹ oder wirtschaftlich motivierten Geschäftsgeheimnissen geschuldet – Umstände, die ihrerseits allerdings als ebenso maßgebliche Quellen technischer Opazität anzusehen sind (vgl. Burrell 2016).[13] Lernfähige Algorithmen werden zur Errechnung und rekursiven Optimierung von Modellen benutzt. Sie werden zur datenbasierten Abbildung je interessierender Zusammenhängen eingesetzt, wobei im Zuge dieser (Selbst-)Programmierung, die gemeinhin »Training« genannt wird, die Maschine entsteht, auf die mit der Formel ML Bezug genommen wird. Genau genommen umfasst die Maschine in ML drei Funktionskomponenten: 1. einen Lernalgorithmus, der eine Funktion zur Beschreibung des zugrundeliegenden Datensatzes errechnet, 2. die errechnete Funktion selbst, auf die ich im Folgenden als »algorithmisches Modell« Bezug nehmen werde sowie 3.

11 Für eine begriffliche Genealogie von »Algorithmus« vgl. Pasquinelli 2019.

12 Entsprechend weist Lev Manovich darauf hin, dass die Rede von »algorithmischer Kultur« irreführend ist, weil sich das damit bezeichnete Phänomen nicht Algorithmen im Allgemeinen, sondern einem spezifischen Typus derselben verdankt, vgl. Manovich 2018, S. 278.

13 Abgesehen davon war technischer Code schon vor der Verbreitung von ML »deep« und daher wesentlich anders zu ›lesen‹ (und verstehen) als ›flach‹ Gedrucktes, weshalb diese Hürde auch nicht als KI/ML-spezifisch angenommen werden könnte, vgl. Hayles 2004.

einen Inferenzalgorithmus, der dem Modell für einen gegebenen Input einen Output entnimmt.

ML-Verfahren entsprechen damit in zweifacher Hinsicht dem, was Heinz von Foerster als »nicht-triviale Maschine« (im Folgenden kurz *NTM*) beschrieben hat (vgl. von Foerster 1993b). Einerseits sei deren Operationslogik insofern als nicht-trivial zu verstehen, dass ihre Outputs durch die Verarbeitung von Inputs mittels »interner Zustände« (ebd., S. 138) erzeugt werden, die für Beobachter:innen unbekannt sind und bleiben, weshalb ihnen die Maschine intransparent und ihre Outputs nicht erwartbar anmuten. Weil sich Beobachterinnen die Beziehung zwischen Inputs und Outputs undurchsichtig zeigt, erweist sich ihnen die Maschine als nicht-trivial. Dies gilt für ML ganz zweifellos. Andererseits trifft von Foersters Beschreibung der NTM auch in dem Sinne auf ML zu, dass beide nicht in erster Linie ein dinghaftes Objekt im Sinne eines unmittelbar beobachtbaren Apparats verweisen, sondern auf ein »conceptual device« (von Foerster 1984, S. 9), an dem primär dessen »process« (ebd.) von Interesse ist – und nicht etwa seine materielle Verfassung. Mit dem nicht-trivialen Charakter von ML geht seine bereits erwähnte Opazität einher. Als unhintergehbare Möglichkeitsbedingung derartiger algorithmischer Modellierungsverfahren ist diese Eigenschaft untrennbar von deren Generativität und Produktivität, wenngleich dies mitunter auch Schwierigkeiten in der Evaluation algorithmischer Modelle bedeutet, was die Operationalisierung von Erfolgskriterien praktischer Anwendungen von ML verkomplizieren kann (vgl. Mackenzie 2017, S. 96). Gegenüber früheren Ansätzen der KI-Forschung scheint dennoch gerade der Verzicht auf die Erklärbarkeit ›intelligenter Maschinen‹ insgesamt den jüngeren Entwicklungen um ML grundlegend den Weg geebnet zu haben. So schreibt Elena Esposito:

> many of the recent successes of digital machines are not because machines have finally learned to understand content, but because programmers have given up trying to produce machines that understand. (Esposito 2021, S. 378)

Operationslogisch lässt sich die maßgeblichen ML-Verfahren eigene Opazität über die Unterschiede ihrer rechentechnischen Architektur im Vergleich zur Von-Neumann-Architektur beschreiben, die das konzeptuelle Fundament des modernen Digitalcomputers darstellt und auf drei wesentlichen Prinzipien beruht. Erstens werden alle Daten digital verarbeitet, zweitens werden Daten, Programme sowie die Rechenergebnisse in gleichen Speicher abgelegt und drittens erfolgt die Verarbeitung von Daten und Programmbefehlen in serieller bzw. sequenzieller Form.[14]

14 Somit stellte sie zugleich alle Funktionskomponenten der Turing-Maschine dar. Diese war von Alan M. Turing im Jahre 1936 als Gedankenexperiment konstruiert worden (vgl. Turing 1936/1937), um zu zeigen, dass jeder mathematisch definierbare Algorithmus mithilfe eines

Sogenannte neuronale Netzwerke – der einflussreichste ML-Ansatz seit 2014 – brechen mit diesen Grundprinzipien digitaler Computer in zweifacher Hinsicht. In diesen von Annahmen über die Funktionsweise des menschlichen Gehirns inspirierten netzwerkförmigen Rechenarchitekturen wird einerseits die netzwerkinterne Gewichtung der Verbindungsstärke zwischen einzelnen Neuronen in der Regel mit Fließkommazahlen vorgenommen, weshalb die internen Verweisstrukturen des Netzes als nicht allein digital verfasst zu verstehen sind.[15] Dies ermögliche so feingliedrige Repräsentationsfunktionen, dass die entsprechenden Werte trotz digitalem Substrat »quasi-analog« (Sudmann 2018b, S. 66) erscheinen. Andererseits ist zu beachten, dass infolge eines Inputs derartig viele Neuronen aktiviert werden (»gemeinsam feuern«), dass

> sie auf diese Weise ein komplexes emergentes System bilden, das letztlich die Diskretheit der Elemente, aus denen es besteht (die Neuronenschichten und ihre Verbindung), fundamental aufhebt […]. (ebd., S. 67)

Es sind dies ebenso die Gründe dafür, dass ML mitunter als »postdigitale Informationstechnologie« (ebd.) beschrieben wird – bzw. als »Unschärfesystem […], dessen Operationen ›eher‹ als analog denn als digital zu beschreiben sind« (ebd.).

Diese medienwissenschaftliche Beurteilung Andreas Sudmanns soll an dieser Stelle nicht ausführlich auf ihre Implikationen für das soziologische Verständnis von ML untersucht werden. Zunächst geht es hier allein um die Feststellung, dass sich infolge der beschriebenen Disposition[16] die Outputs künstlicher neuronaler Netze nicht plausibel genetisch rekonstruieren lassen, weil, wie es der Informatiker Roland Memisevic im Interview mit Sudmann formuliert: »emergente Phänomene auftauchen, wo das Ganze mehr als die Summe seiner Teile darstellt« (Memisevic & Sudmann 2018, S. 377f.). Dies hat zur Folge, dass bei allen derartig ›tiefen‹ ML-Verfahren mit nicht-linearer (und nicht-trivialer) Funktionsweise eine prinzipielle

---

einfachen Apparats prinzipiell ausführbar ist. Turing schlägt als einen solchen eine Maschine vor, die auf eine endlose Papierrolle schreibt und das Geschriebene (an beliebiger Stelle) lesen wie auch löschen kann. Auf dieser Basis könne sie jede andere korrekt programmierte Maschine simulieren. Friedrich Kittler zufolge habe Turing mit diesem Entwurf im Grunde den konzeptuellen Schlussstein einer medientechnischen Entwicklungsgeschichte gesetzt: »[T]he Turing machine concluded in its universality all developments for the storing, indexing, and processing of both alphabetical and numerical data« (Kittler 1996, S. 11).

15 »Nicht allein«, weil sie dennoch zugleich mit der Unterscheidung aktiv/passiv beschrieben werden können (vgl. Sudmann 2018b, S. 66).

16 Die übrigens ebenso zur Folge hat, dass einzelnen Ereignissen eines neuronalen Netzwerks keine eindeutige Adresse zugewiesen werden kann – auch dies ein wesentlicher Unterschied zu ›klassischen‹ Digitalcomputern.

operative Opazität vorherrscht, die mit beschränkter Interpretierbarkeit (auch post-hoc) i.S.v. Nachvollziehbarkeit von Outputs der Anwendung einhergeht.[17]

Es bedarf dann kaum einer gesonderten Erörterung, dass sich dieses Problem in der Nutzung ›zweiter Hand‹ nicht etwa aufhebt, sondern fortsetzt. So bringt die Weiterverwendung von unter Bedingungen rechnerischer Opazität entstandener Daten ebensolche Unwägbarkeiten mit sich. Da zu vielen kommunikativen Anlässen der fragile epistemische Status von durch ML produzierten Datenanalysen entweder nicht explizit thematisiert wird und/oder die Quellen von als situativ relevantem Wissen Beobachteten schlichtweg unbekannt sind, kann sich ›Desinformation‹ daher auch ›hinter dem Rücken‹ der Betroffenen vollziehen. Diese Diagnose gilt durchaus umfassend, weil davon ausgehen zu ist, dass in praktischen sozialen Zusammenhänge keine epistemologische Dauerreflexion mit Blick auf situativ relevante Wissensquellen stattfindet. Demzufolge hat man es in soziologischer Betrachtung bei jedem praktischen Einsatz von ML-Verfahren genau genommen gewissermaßen mit »multipler Opazität« (Roberge & Seyfert 2017, S. 9) zu tun, weil die ursprüngliche Implikation alle an empirisch anschließenden »Relationierungen innerhalb einer Fülle von menschlichen und nicht-menschlichen Akteuren« (ebd.) erfasst. Es ist dieser epistemologische Aspekt rechentechnischer Intransparenz, der ML von anderen Formen digitaler Rechentechnik maßgeblich unterscheidet.

Besonders gut lässt sich diese These am Beispiel solcher heutzutage ML-basierter Softwaredienste illustrieren, die vormals eine andere rechentechnische Grundlage hatten. Die Oberfläche von Websites kann im Zuge von Umstellungen in der technischen Infrastruktur der angebotenen Dienste gleichgeblieben sein – man denke etwa an alle wesentlichen Social Media-Plattformen oder die Google-Suchfunktion[18] – doch kann sich die Grundlage der Errechnung von Inhalten wie Suchergebnissen und damit die Funktionsweise des Dienstes maßgeblich verändert haben, ohne dass User:innen die erfolgte Umstellung notwendigerweise registriert haben müssten. In diesem Sinne ist ›der Algorithmus‹ zum geflügelten Wort und mithin zur allgemeinen Chiffre für Überraschungserfahrungen im Kontext von *UX* (*User Experience*) geworden, für die betroffene Nutzer:innen keine Erklärung finden können. Folglich haftet dem Begriff eine semantische Unschärfe an, die an das *moving target* KI/ML erinnert, das aufgrund seiner latenten Unbestimmtheit seit jeher als Projektionsfläche für vielerlei Imaginäre dient. ›Dem Algorithmus‹ oder ›der KI‹

17 Dabei kann indes weiter differenziert werden zwischen der Opazität des gesamten Modells, individueller Komponenten und des Trainingsalgorithmus. Diese Ebenendifferenzierung betrifft mit Blick auf Aspekte von Interpretierbarkeit (in gleicher Reihenfolge) die Simulierbarkeit und (analytische) Zerlegbarkeit des Modells sowie die operative Transparenz des Algorithmus, vgl. Lipton 2018.

18 Vgl. für deren Pionierrolle im Zuge des »Great AI Awakening« (Lewis-Kraus 2016) vgl. Kap. 1, S. 5f.

kann vieles zugeschrieben werden und gleichermaßen wird beiden als imaginierten Agenten auch realiter eine Vielzahl von Eigenschaften, wenn nicht gar Absichten und mithin auch Verantwortung zugeschrieben. Nicht selten geschieht dies gerade auch dann, wenn die Betroffenen keine plausibel rationalisierende Erklärung für einen gegebenen Umstand finden. Deshalb ist für KI, ML wie auch für Algorithmen besonders hinsichtlich methodologischer Überlegungen unbedingt zu beachten, dass alle drei im praktischen Gebrauch (als in-vivo Codes) keinesfalls über eine eindeutige empirische Referenz verfügen.[19]

In diesem Zusammenhang lässt sich ausgehenden von dieser Feststellung einer gewissen Beliebigkeit eine interessante semantischen Verschiebung beobachten, die sich auf die Beschreibung der ›Dingwelt‹ bezieht und einen Wandel von Registern des Passiven bzw. Widerständigen in das Aktive und Kognitive suggeriert. ›Wollte‹ der Nagel nicht in die Wand, ›streikte‹ der Drucker häufig und ›hing‹ sich der Computer hin und wieder ›auf‹, scheint es sich bei Software und anderen primär als programmiert wahrgenommenen und gewissermaßen immateriell erlebten Artefakten auf prägnante Weise anders zu verhalten. ›Der Algorithmus‹ (und gleichermaßen: ›die KI‹) ›entscheidet‹, ›analysiert‹, ›schlägt etwas vor‹ oder ›hat etwas entdeckt‹.

Einen hilfreiche Erklärungsansatz für diese Tendenz wie auch für jene semantische Unschärfe mit Blick ›Algorithmen‹ bietet Frieder Nakes Unterscheidung von »Surface« und »Subface« (vgl. Nake 2008). Mit dieser Differenzierung von – durch das »Interface« verbundenen – Betrachtungsebenen lässt sich nachvollziehbar markieren, worin ein in technischer Hinsicht entscheidender Unterschied zwischen ML und ›herkömmlichen‹ Digitalcomputern[20] besteht. Dieser lässt sich im veränderten Verhältnis erkennen, in das nicht nur die Nutzer:innen, sondern auch die für die Programmierung Verantwortlichen zur ›Unterfläche‹ (Subface) der Technik geraten, der elementaren Basisebene ihrer Operationen, auf der die Programmierung der Oberfläche stattfindet, wo entsprechenden Algorithmen arrangiert und angepasst werden. Die beschriebene Opazität von für ML verwendete Algorithmen sorgt

19 Empirisch lässt sich dann freilich nicht ohne Weiteres unterscheiden, ob zur Erklärung entsprechender Zuschreibungen im konkreten Fall entsprechend Jenna Burrells Opazitätsanalytik (vgl. Burrell 2016) ›technisches Analphabetentum‹, per Design bewusst herbeigeführte Intransparenz oder die ML eigene rechentechnische Opazität (oder aber eine Kombination der drei) herangezogen werden sollte.

20 Etwa im grundlegend Sinne einer programmierbaren von Neumann-Maschine. Diese Differenzierungslinien ließe sich weiter verkomplizieren, wollte man dem Umstand Rechnung tragen, dass die Berechnung (*computation*) eines neuronalen Netzes auf solchen ›herkömmlichen‹ Maschinen ausgeführt wird (wenn auch spezifisch ausgerüsteten, etwa mit besonders schnell parallel rechnenden Grafikprozessoren), ML also ›normale‹ Computer als Recheninfrastruktur in Anspruch nimmt.

für eine Disposition, in der sich diese Ebene eines Computers dem direkten Zugriff völlig entzieht. Dieser Umstand könnte beschrieben werden als eine Verschiebung von Operationsebenen: das tatsächliche *subface* ist nicht unmittelbar zugänglich und alle Beteiligten, auch die Konstrukteur:innen der Maschine, können diese allein über *interfaces* bedienen.[21] Folglich erscheint die ›Basis-Programmierung‹ *allen* Beteiligten prinzipiell als alternativlos opak. Es zeigt sich hier eine Disposition, in der sich Algorithmen als elementare Bestandteile von ML grundlegend nicht beobachten lassen und aber zugleich – wie die erwähnte semantische Verschiebung suggeriert – mit zunehmend ›anspruchsvolleren Aufgaben betraut werden‹.

Diese saloppe Formulierung verweist auf ein Phänomen, das Florian Jaton in seiner umfassenden Studie zur Genese von Algorithmen analytisch präziser als »Delegation« beschreibt (vgl. Jaton 2021, S. 278): die für ML typische automatische Errechnung einer Funktion (Modellierung) durch den Einsatz von Lernalgorithmen leistet die Formulierung eines Modells auf Kosten der genauen Kenntnis desselben.[22] Die Aufgabe der Formulierung wird in diesem Sinne an den Algorithmus delegiert, wobei zu beachten ist, dass dieses Arrangement in aller Regel praktisch eine Aufschichtung (»stacking«) verschiedener Algorithmen bedeutet, die sozusagen als Infrastruktur der Modellierung fungiert:

> the more machine learning, the more delegation, and the more difficult it becomes to inspect what has led to the formation of the mathematical operation allowing the transformation of inputs into outputs. (ebd.)

Ausgehend von diesem Unterschied zwischen ML und ›klassischer‹ Programmierung erscheint eine ganze Reihe von Fragestellungen bezüglich »algorithmischer Kulturen« gegenüber dem Stand von vor fast 20 Jahren (vgl. etwa Galloway 2006) tatsächlich in einem neuen Licht.[23]

> Was tun Algorithmen, was bringen sie kulturell hervor? Wie generieren sie Sinn aus ihren Umgebungen und den verschiedenen Kategorien, die Menschen nutzen, um die Algorithmen zu deuten? (Roberge & Seyfert 2017, S. 14)

Ausgehend von der Beobachtung von Delegation (Jaton), der Selbstprogrammierung von Computern mittels Lernalgorithmen, erscheinen derartige Fragen

21 Womit ein wesentlicher Ansatzpunkt für Ansätze von sogenannter »Explainable AI« markiert ist.

22 »Algorithmic Modeling« unterscheidet sich von anderen Formen des »data modeling« (vgl. Breiman 2001) folglich dadurch, dass die Modellierung sich selbst strukturiert, ohne dass Modellvariablen (Faktoren und Parameter) a priori festgelegt werden – dies wird im beschriebenen Sinne an den Algorithmus delegiert.

23 Gleiches betrifft die semantische Assoziation von Algorithmen und KI.

durchaus empirisch plausibel und wissenschaftlich legitim. Jonathan Roberge & Robert Seyfert suggerierten indes schon 2017, dass »Algorithmen Teil der Sinnstiftung des kulturellen Imaginären geworden sind« (ebd., S. 16), etwa angesichts von Praktiken wie des ›Googlens‹, das maßgeblich auf den PageRank-Algorithmus baut. Solche Arrangements, die »symbolische und performative Aspekte stetig interagieren« (ebd., S. 15) lassen mit algorithmischen Berechnungen, machen es naheliegend, dass Algorithmen tatsächliche in der sozialen Wirklichkeit eine eigene Art von Wirkung »entfalten« (Lee et al. 2019, Übers. RG).[24]

Es soll an dieser Stelle für die weitere Entwicklung des Arguments festgehalten werden, dass jenes für den Gegenstand dieser Studie namensgebende ›Lernen‹ in einem engen Zusammenhang mit der Rolle von spezifischen adaptiven Algorithmen steht, die Datensätze rekursiv rechnerisch in Modelle überführen. Diese Assoziation ist allerdings noch keine Erklärung der verantwortlichen Mechanismen für dieses technische Verständnis von ›Lernen‹, wie es sich im Zusammenhang mit ML etabliert hat. Die weitere Erörterung dieser Verknüpfung geht zunächst einen historischen Umweg, weil das ML und mithin ›KI‹ umgebende semantische Assoziationsfeld nicht bei Verständnissen des Lernens stehen bleibt, sondern ebenso die Topoi der »Intelligenz« sowie der »Künstlichkeit« berührt.

ML & GOFAI

In *Good Old Fashioned ArtificiaI Intelligence (GOFAI)* – etwa durch die sog. Expertensysteme der 1990er Jahre bekannt[25] – wird eine (von menschlichen Expert:innen geschriebene) Programmierung als Quelle dessen aufgefasst, was Beobachter:innen in den Resultaten der Anwendung des Programms intelligent erscheinen mag. Klassischerweise wird dieser, auch als symbolische KI bekannte Ansatz des »making the mind« vom subsymbolisch (bzw. konnexionistisch) genannten Paradigma des »modelling the brain« unterschieden, dem ML zugerechnet wird (vgl. Dreyfus/Dreyfus 1988). Ein von Ingenieur:innen entwickeltes, möglichst elaboriertes Regelsystem (algorithmisch im Sinne eines »well-defined computational procedure« (Burke 2019, S. 1) wird als Software implementiert, was dieses zur (einzigen) kognitiven Grundlage der ›Expertise des Expertensystems‹ macht.[26]

---

24 Die gerade sozialtheoretisch relevante und im Kontext dieser Untersuchung auch methodologisch relevante Frage, ob die »Entfaltung« der Effekte des Einsatzes von Algorithmen dabei – wie im Zitat – als Interaktion treffend beschrieben ist oder besser anders gefasst werden sollte, klammere ich an dieser Stelle zunächst ein. Weiter unten werde ich sie erneut aufzugreifen, wenn es um die Beschreibung von Sozialität unter Beteiligung von ML gehen wird.

25 Die Beiträge des einschlägigen Bandes »Soziologie und Künstliche Intelligenz« aus dem Jahre 1995 etwa haben wesentlich Expertensysteme zum Gegenstand (vgl. Rammert [Hg.] 1995).

26 Dies gilt freilich nur für die Betrachtung des Expertensystems als Technik für sich. Soziologisch betrachtet ist Expertenwissen demgegenüber vielmehr »weder in der Maschine noch

Üblicherweise enthält *GOFAI* keine adaptive Komponente[27], weshalb sich mit diesem Ansatz, der die KI-Debatten über Jahrzehnte prägte, kaum Assoziationen zu Konzepten des Lernens aufbauen lassen, wie sie für ML charakteristisch sind. Eine konstitutive Limitierung für diese Variante von *GOFAI* besteht darin, dass ihre Programmierung präskriptiv determiniert ist und ›Lernen‹ im Sinne einer Anpassung der operativen Logik und Struktur nur durch externe Interventionen des zuständigen Ingenieursteams möglich ist. Die aus der Beobachtung der Maschine gewonnenen Einsichten kann das Team prinzipiell auch zu späteren Zeitpunkten in die (Weiter-)Programmierung der Maschine einfließen lassen. Dies geschieht additiv in Form neuer ›Häppchen‹ von gewissermaßen in Rezeptform in die Maschine transferiertem Wissen, oder subtraktiv durch das Entfernen bestehender Programmelemente. Insofern in diesem Zusammenhang von Lernen die Rede ist, wäre dies folglich dem ›Programmierteam‹ zuzurechnen; die Maschine für sich weist keine adaptiven Mechanismen auf.

Die ›Intelligenz‹ der ›KI‹ im Rahmen von *GOFAI* konnte daher konstruktionsbedingt nicht die derjenigen übertreffen, die mit der Programmierung zugleich die Grenzen des Variantenreichtums der maschinellen Berechnungen determinierten. Die *top down*-Logik des Ansatzes sah vor, dass die Maschine aus festgelegten Strukturen heraus alles Relevante gewissermaßen deduzieren können sollte, weshalb in Form der Programmierung im Grunde bereits vorab alles operativ Relevante algorithmisch expliziert sein musste (vgl. Sudmann 2018b, S. 59). Folglich sind strukturbedingt keine internen (rechentechnischen) Emergenzeffekte denkbar[28] – das Informationspotenzial der KI beschränkt sich auf das per Programmierung abgelegte Wissen. Kybernetisch formuliert ist *GOFAI* mithin gerade nicht zur Errechnung von »Eigenwerten« (von Foerster 1993c, S. 107) fähig, weil die möglichen »internen Zustände« (von Foerster 1993b, S. 138) einer derartigen Maschine von vornherein bekannt sind, da ihre Funktion expliziert programmiert wurde. Mit einer überraschende Outputs zeitigenden *black box* hätte man es deshalb nur dann zu tun, wenn die Funktionsweise (Programmierung) der Maschine nicht vollständig bekannt ist.

Gegenüber *GOFAI* erscheint die *Selbst*programmierung mittels automatischer Datenverarbeitung als funktionaler Kern maschineller Lernverfahren. Dieser zen-

---

im Menschen existent, sondern es entsteht erst in ihrer Wechselwirkung. Es findet seine effektive Form zunächst in der ›experimentellen Interaktivitat‹ und später in der ›routinisierten Interaktivitat‹ zwischen Nutzer und Programm« (Rammert 2007, S. 160), ist mithin also weniger eine statische, auf Knopfdruck verfügbares Gut, sondern in der Praxis eine dynamische Ressource, deren Erschließung nicht-linear erfolgt.

27 Wenngleich gegenwärtig gelegentlich Plädoyers für hybride Varianten von KI, die vermeintlich die Schwächen beider Paradigmen zu überwinden imstande seien, diskutiert werden, vgl. etwa Marcus 2020.

28 Was freilich nicht ausschließt, dass der praktische Einsatz einer solchen Maschine dennoch praktisch für Überraschungen sorgt.

trale Aspekt der Errechnung einer Funktion zur Beschreibung eines Datensatzes[29] wird in der KI-Forschung als Training bezeichnet. Ist von »Künstlicher Intelligenz« die Rede, bezieht sich diese Formel folglich direkt auf das ›im Training Gelernte‹, ob als manifester Output oder in latenter Form der Funktion, die Outputs determiniert.[30] Automatisierte Datenanalyse im Sinne der Errechnung einer Funktion zur Beschreibung Beziehungen von Datenpunkten als Selbstprogrammierung einer Maschine markiert daher den operationslogischen Kern von ML und mithin jenen rechentechnischen Kognitionsvorgang, der gemeinhin Mustererkennung genannt wird.

GOFAI gegenüber wird in ML das operative Funktionsprinzip im Grunde umgekehrt: statt von aus festgelegten Strukturen (Programmierung) in *top down*-Manier Inputs zu verrechnen, programmiert sich die Maschine qua algorithmischer Datenanalyse selbst: operationslogisch betrachtet geht es *bottom up* von vorgegebenen Daten (›Beispielen‹) zur Programmierung.[31] Die Möglichkeit einer solchen de facto induktiven Selbstprogrammierung verdankt sich adaptiven Algorithmen, die im Zuge des ›Trainings‹ eine sukzessive Optimierung der Modellierung des Datensatzes erreichen.[32] Dessen Sinn, d.h. das aus der Perspektive von Beobachter:innen im Datensatz befindliche Wissen muss dabei folglich als mathematisch beschreibbar angenommen werden. Schließlich dient die errechnete Funktion in der Folge als Modell für die Produktion weiterer Daten. Jeder neue Input kann dann im Vergleich mit dem Modell (das daher die »ground truth« ›des Algorithmus‹ darstellt) hinsicht-

29 Was in epistemologischer Reflexion impliziert, dass eine solche Funktion, aus der die Daten hervorgegangen sind, sich prinzipiell errechnen lässt und die Datenbasis als *ground truth* prinzipiell durch eine Funktion beschrieben werden kann, vgl. Mackenzie 2017, S. 82.

30 Wenngleich dies nicht heißt, dass Expert:innen keinen Anteil an der praktischen Genese einer ML-Anwendungen hätten. In soziologischer Perspektive auf ML und *GOFAI* ist deshalb von einer allzu vereinfachenden Gegenüberstellung beider Ansätze abzusehen, besonders hinsichtlich der Zurechnung von Verantwortung für die Effekte einer Anwendung. Gegenüber unrealistischen Suggestionen bezüglich der ›autonomen Lernfähigkeiten‹ algorithmischer Modellierungen kann nüchtern konstatiert werden: ML-Anwendungen »often work well because structures, priors, representations, invariants and background knowledge were designed into the machine by a domain expert« (Jebara 2004, S. 10) – und diese Aussage wäre auch für Expertensysteme zutreffend.

31 Damit geht freilich einher, dass die Funktionalität eines algorithmischen Modells grundlegend von der Qualität des Datensatzes abhängt, das diesem als *ground truth* dient. Insofern ließe sich argumentieren, dass ›Expertenwissen‹ (und Fähigkeiten) wie in *GOFAI* jedenfalls in der Erhebung und/oder Zusammenstellung sowie Aufbereitung der Datengrundlage der Modellierung eine entscheidende Rolle spielen, denn: » […] we get the algorithms of our ground truths« (Jaton 2021, S. 24).

32 Die Abweichung der Modellierungsfunktion von den Trainingsdaten wird dabei als »loss function« angegeben, vgl. Mackenzie 2017, S. 96.

lich seiner Eigenschaften analysiert oder – im Falle generativer Verfahren – dazu genutzt werden, Daten zu erzeugen, die zum Datensatz ›gehört haben könnten‹.

Bereits in diesen rudimentären Ausführungen sollte nachvollziehbar geworden sein, dass ML gegenüber *GOFAI* rechentechnisch betrachtet fundamental anders operiert. Insofern könnte man sich darüber wundern, dass dennoch beide gemeinhin als Formen *künstlicher Intelligenz* subsumiert werden. Die Operationslogik von *GOFAI* kann intuitiv leichter nachvollzogen werden, da sie wesentlich der konventionellen Vorstellung eines ›klassischen‹ Computerprogramms entspricht. Dieses wird von Programmierer:innen formuliert und kann dann schließlich von Endnutzer:innen noch final konfiguriert werden. Die Herstellung des Programms kann dabei eindeutig denjenigen zugeschrieben werden, die es formuliert haben. Bei ML verhält es sich hingegen vertrackter, wird hier doch ein algorithmisches Verfahren genutzt, um die Formulierung des Programms zu automatisieren, wodurch diese Aufgabe gewissermaßen an den Algorithmus delegiert wird (vgl. Jaton 2021, S. 278). In diesem Arrangement ist es daher das algorithmische Verfahren selbst, das in Kombination mit den Trainingsdaten sozusagen verantwortlich für die finale Funktion, das Modell, den Algorithmus ›in trainierter Form‹ ist.[33]

Doch nicht nur hinsichtlich ihrer technischen Operationslogik, sondern ebenso mit Blick auf den ihnen impliziten Wirklichkeitsbezug unterscheiden sich ML und *GOFAI* maßgeblich. Dies zeigt sich auf prägnante Weise in den mit den jeweiligen Ansätzen typischerweise assoziierten Metaphern. So charakterisieren, wie oben bereits erwähnt Hubert und Stuart Dreyfus symbolische KI als den Versuch, auf technischem Wege »Geistigkeit zu erzeugen« (*making a mind*), wohingegen ML als subsymbolisch-neuronale KI dem Anliegen folge, »das Gehirn zu modellieren« (*modelling the brain*) (Dreyfus & Dreyfus 1988). Diese Vergleiche mögen in mancher Hinsicht schief sein – sowohl bezüglich der ›Geistigkeit‹ von *GOFAI* als auch des Anliegens subsymbolischer KI. So wie Geistigkeit kaum allein in der logisch konsistenten Anwendung von Regeln besteht, muss ebenso als fraglich gelten, ob die Aktivitäten des menschlichen Gehirns[34] als algorithmisch errechnete Funktion adäquat beschrieben sind.

---

33 Für jede Analyse von *agency* (und deren Konstitution) beschert dies offensichtlich eine komplizierte Gemengelage in diesem Zusammenhang. Will man für einen beliebigen Output nach dessen genetischer Struktur fragen, kommen unmittelbar mehrere Instanzen in den Sinn, denn weder ›der Algorithmus‹ noch ›die Daten‹ sind ursprüngliche Quellen von Sozialität, verweisen sie doch wiederum auf Vorgängiges – etwa auf die Entwicklung des algorithmischen Verfahrens sowie dessen Auswahl für die betreffende Anwendung oder auf das empirische Referenzobjekt der Daten sowie Erhebungs- und Kuratierungsaktivitäten auf deren ›Weg in den Datensatz‹.

34 Und vor allem deren emergente Effekte, man denke gerade hier an ›Geistigkeit‹ in Form menschlichen Bewusstseins.

Der Verweis auf das menschliche Gehirn dient in subsymbolischer KI vor allem als Inspiration für den Entwurf von Architekturen und operativen Dispositionen für die die Speicherung und Verarbeitung von Information (im rechentechnischen Sinne der Verarbeitung reiner Differenz: ›0/1‹). Folglich ist mit Blick auf ML wie auch *GOFAI* nicht unmittelbar ersichtlich, auf welcher Basis ein »ghost in the machine« zu vermuten wäre, was insbesondere bei *GOFAI* angesichts des mehr oder minder explizit derart gesetzten Ziels verwunderlich erscheinen kann. »Making the mind« scheint gewissermaßen mit »making a set of rules« gleichgesetzt, ohne dass Geistigkeit in ihren Eigenheiten thematisch würde. Besteht eine klassische Ansicht von Hegels Philosophie des Geistes darin, dass (subjektiver) Geist im Bewusstsein »für sich vermittelt« ist[35], wäre etwa mit Blick auf KI (gleich welcher Couleur) zu fragen, was eigentlich als das (ihr eigene) Medium ihrer Operationen gelten kann. Mit Blick auf die reale rechentechnische Operationsbasis (»0/1«) sowohl von symbolischer wie subsymbolischer KI scheint Information als reine Differenz (vgl. Bateson 1983b, S. 353) einen geeigneten Kandidaten darzustellen. In dieser Hinsicht scheint das Gehirn als Metapher für Computertechnik überhaupt ungeachtet aller weiteren Implikationen zumindest weniger irreführend zu sein.

Dies liegt daran, dass »Informationsverarbeitung« (im vorgestellten Sinne) deutlich niedrigschwelliger und mithin technisch einfacher zu realisieren scheint als etwa »Geist« oder »Bewusstsein«.[36] Die Annahme einer Äquivalenz von KI und Geist mag grundsätzlich einen Kategorienfehler konstituieren – in soziologischer Betrachtung jedenfalls – doch muss dies hier nicht weiter in den Blick genommen werden, weil für das Argument alleine die die Seite der technischen Informationsverarbeitung zum Zwecke der Modellierung von sozialer Wirklichkeit wie im Falle von ML von Belang ist. Dies gilt hinsichtlich ihrer inhärenten impliziten Sozialität – wie wird soziale Wirklichkeit im Modell abgebildet? – wie auch mit Blick auf die realen Folgen des Einsatzes derartig informationsverarbeitender Technik. Bezüglich ersterer Dimension ist von grundlegendem Interesse, wie überhaupt die Rede davon sein kann, dass Informationsverarbeitung in rechentechnischen Verfahren auf sinnförmig verfasste Sozialität bezogen ist. Am Beispiel von Kommunikation möchte ich diesen Zusammenhang im Folgenden technikhistorisch erörtern.

35 Der subjektive Geist ist hier »an sich oder unmittelbar«, wohingegen er im Bewusstsein als »für sich vermittelt« und im Geist als »sich in sich bestimmend« erscheint (vgl. Hegel 1970 [1817], S. 38 [E III 38]).

36 Auf dieser argumentativen Linie erschienen – allen Prätentionen zum Trotz – geistreichen anmutende Formen von KI niemals als echte Instanzen von Geistigkeit oder Wahrnehmung, doch gegebenenfalls immerhin als gelungen Simulationen derselben.

Die Dualität von Information und Kommunikation

Der Nexus von Information und sinnförmiger Sozialität wie Kommunikation, wie er für ML maßgeblich ist, weist technikhistorisch mindestens 75 Jahre zurück und bildet gewissermaßen das Fundament der ›modernen‹ Kommunikationstheorie. 1949 erschien nämlich Claude Shannons (zuerst 1948 als zweiteiliger Aufsatz publizierte [vgl. Shannon 1948]) »Mathematical Theory of Communication«, ergänzt um einen vorangestellten Aufsatz von Warren Weaver in Buchform (vgl. Shannon/Weaver 1949).

Shannon beschreibt in seinem Aufsatz Kommunikation als technisch-statistisches Phänomen, bestimmt durch das Verhältnis von Rauschen und Signal in einer Nachricht. In seiner Beschäftigung als Ingenieur des ›Kommunikationsunternehmens‹ AT&T bestand eine wesentliche Aufgabe Shannons darin, mit technischen Mitteln gestörte (›rauschende‹) Nachrichtenübermittlung – etwa in einer Telefonleitung – derart zu korrigieren, dass der Signalanteil und mithin die Wahrscheinlichkeit dafür steigt, dass eine gesendete Nachricht wie gewünscht empfangen wird. Die Unterscheidung von Signal und Rauschen bezieht sich bei Shannon nicht auf semantische Register oder den Sinn einer Nachricht, sondern auf die quantifizierbare Qualität der Verbindung, die deren Übermittlung ermöglicht.

Es war der Mathematiker Weaver – zu der Zeit als Direktor der naturwissenschaftlichen Abteilung der Rockefeller Foundation auf vielseitige Weise um die Förderung interdisziplinärer Forschung auf Basis damals neuartiger informationswissenschaftlicher Perspektiven bemüht –, der in seinem Aufsatz eine duale Perspektive auf Kommunikation entwickelte (vgl. Weaver 1949, siehe auch Geoghegan 2022, S. 22). Dieser zufolge stelle sich für Kommunikation nicht nur das von Shannon bearbeitete »technical problem« der Nachrichtenübermittelung (»How accurately can the symbols of communication be transmitted? [Weaver 1949, S. 4]), sondern ebenso das, was Weaver als semantische Probleme beschreibt:

The semantic problems are concerned with the identity, or satisfactorily cIose approximation, in the interpretation of meaning by the receiver, as compared with the intended meaning of the sender. (ebd.)

Weaver beschreibt hier, was der Soziologie seit Alfred Schütz' Überlegungen zu Max Webers Begriff der Sinnadäquanz (vgl. Schütz 1932, S. 267ff.) als Problemaspekt des Sinnverstehens bekannt ist, und bezeichnet diesen Zusammenhang gegenüber dem technischen »LEVEL A« von Kommunikation als deren »LEVEL B« (Weaver 1949, S. 4).[37] Sein Differenzierungsschema impliziert, dass Information

37 Interessanterweise identifiziert Weaver darüber hinaus eine weitere soziologisch zentrale, für das Argument an dieser Stelle jedoch nicht relevante Ebene von Kommunikation, die er »LEVEL C« bzw. »the effectiveness problem« nennt und die sich auf den Zusammenhang zwischen beabsichtigten und tatsächlichen Kommunikationsfolgen bezieht: »How effectively

als Element von Kommunikation nicht nur als berechenbares elektrisches Signal (Shannon bzw. »LEVEL A«), sondern (auch) sinnförmig verstanden wird (»meaning – »LEVEL B«).

Diese duale Konzeptualisierung von Kommunikation war wissenschafts- wie technikgeschichtlich überaus folgenreich, weil sie für die weitere Entwicklung ›informationswissenschaftlichen‹ Denkens die Identifikation von Information mit ›gespeichertem Wissen‹ (»stored knowledge«) ermöglichte und die Informatik damit von der Beschäftigung mit rein technischen Aspekten der Signalübertragung zu emanzipieren half (vgl. Rieder 2017, S. 106). Fortan ließen sich Sinnverarbeitung (etwa mittels Interaktion) und technische Datenverarbeitung gleichermaßen als Informationsverarbeitende Prozesse plausibilisieren, was zunächst vor allem zu vermeintlich nicht allzu aufregender Arbeit an der Technisierung von Dokumentenauswahl- und Sortierungsverfahren geführt haben möge (vgl. ebd.). Doch waren dies nur die ersten Etappen einer Entwicklung, die sich etwa im Zuge des Aufstiegs der Kognitionswissenschaften bald auch als grundlegende epistemische Verschiebung artikulieren würde. Diese Entwicklung bracht den Aufbau eines umfassenden »imaginary of the computational sensorioum« (Suchman 2021, S. 70) mit sich, das neue Formen der gesellschaftlichen (und menschlichen) Selbstbeschreibung ermöglichte.

Grundlage dafür ist die beschriebene duale Perspektive auf Kommunikation, die über einen technologisch wie phänomenologisch plausiblen, auf Differenz bauenden Informationsbegriff integriert ist.[38] Auf der einen Seite der Perspektive betrachtet Shannon Sprache »nicht als System von Bedeutungen, sondern als System von Wahrscheinlichkeiten« (Müller 2015, S. 22), was einem »*postsignfikatorischen* Denken des Sinns« (ebd., S. 23, Hervorh. i.O.)[39] gleichkommt, demzufolge besonders informativ sein kann, was unwahrscheinlich ist (bzw. für Varianz sorgt), während hohe Redundanz für wenig Information steht.[40]

---

does the received meaning affect conduct in the desired way?« (ebd.). Damit scheint er bereits als Differenz (von B und C) als vor Augen zu haben, was Luhmann später als Unterscheidung der Selektionen von Information und Mitteilung beschreibt.

38 Ein solcher wäre Batesons des ›unterschiedsmachenden Unterschied‹ (vgl. Bateson 1983b, S. 353).

39 Auch Erich Hörl (2011, S. 9) geht in seiner These der technologischen Sinnverschiebung von diesem historischen Punkt als epochalem Ereignis aus.

40 Diese Konzeption entstand vor dem vermutlich nicht unwesentlichen historischen Hintergrund, dass Shannon im Kontext des Zweiten Weltkriegs als Kryptoanalyst für das US-amerikanische Militär tätig war. Und nicht gänzlich überraschend ging der Publikation seiner »Mathematical Theory of Communication« (1948) im Jahre 1945 die einer »Mathematical Theory of Cryptography« aus seiner Feder voraus. Wohl weil in der militärischen Verschlüsselungstechnik die Gleichwahrscheinlichkeit aller verschlüsselten Symbole als höchstes Ziel erscheint, weht in Shannons Informationstheorie gewissermaßen auch der Geist einer Theorie der *Des*information (vgl. Müller 2015, S. 22).

Manche Verständnisschwierigkeiten im Umgang mit ML erscheinen im Lichte dieser historischen Entwicklungen als späte Folgen einer epistemischen Verschiebung, die in den 1940er Jahren ihren Anfang nahm. Dass im Zuge der Errechnung wahrscheinlicher Sinngehalte in ML mitunter Wider- und Unsinniges technisch produziert wird, erweist sich so umso deutlicher nicht etwa als Manifestation eines ›ghost in the machine‹, sondern stattdessen viel eher als ›Unfall‹ im Versuch der Überwindung einer durch die Inkommensurabilität zweier Verständnisse von Information bedingten epistemischen Lücke. Wie kann im Lichte dieser Beobachtungen eine instruktive Perspektive auf Informationsverarbeitung mittels algorithmischer Operationen, wie sie für ML typisch sind, konstruiert und insbesondere der ML auszeichnende Aspekt des *Lernens* soziologisch verstanden werden?

Information und Lernen

Für die Technik- und Medienphilosophin Luciana Parisi steht die für ML zentrale »algorithmische Entdeckung von Information« als epistemologische wie technologische Errungenschaft in engem Zusammenhang mit Vorstellungen von »selbstadaptivem Verhalten« und der »Mechanisierung des logischen Denkens« (vgl. Parisi 2018, S. 93f.), wie sie für die Kybernetik prägend waren. Die über adaptive Algorithmen vermittelte maschinelle ›Lernfähigkeit‹ zeichne sich vor allem dadurch aus, dass sie gegenüber früheren Verfahren Zufall nicht mehr zu vermeiden suchen müsse, sondern etwa im Sinne von *trial and error* verarbeiten könne (vgl. ebd.). Diese Algorithmen ermöglichten mithin »adaptive Lernverfahren, die dazu fähig sind, Muster zu erkennen oder vorherzusagen, ohne dass sie darauf programmiert worden wären oder das Ergebnis ihrer Aktion im Vorhinein kennen würden« (ebd., S. 93). Der kybernetische Fokus auf Feedbackmechanismen, etwa in Modellen wie Claude Shannons »Theseus« oder Ross W. Ashbys »Homöostat«, hätte zur Entstehung von Vorstellungen dazu geführt, wie Maschinen das »Lernen lernen« könnten (vgl. ebd., S. 95). Diese hätten weggeführt von der Emphase auf die ›Vor-Programmierung‹ (und mithin Prädeterminierung) von Maschinen und zugleich den Weg geebnet für die semantische Parallelisierung von Denken und Rechnen.[41]

Im Kontext dieser soziologischen Studie interessiert allerdings weniger der Zusammenhang zwischen lernenden Maschinen und Denken, sondern vielmehr das Verhältnis der Operationen dieser Maschinen zu sozialer ›Sinnverarbeitung‹, d.h. etwa der Beobachtung dieser Operationen oder ihrer Beteiligung an Kommunikation. Parisis Reflexionen stehen jedoch in einem engen Zusammenhang mit den eben vorgestellten Überlegungen Weavers zum Informationsbegriff, in denen mittels der

41 Vgl. weiterführend dazu soziologisch Nassehi 2019, S. 235ff. sowie philosophisch Günther 2021 [1953].

Erweiterung von Shannons Kommunikationsbegriff gleichermaßen die Möglichkeit einer Parallelisierung von technischen und semantischen Aspekten von Information erkannt wird. Weaver wie Parisi markieren eine für den sozialen Einsatz von adaptiver Rechentechnik konstitutive soziale Disposition, die sich womöglich gegenwärtig unter dem Eindruck der zunehmenden sozialen Verbreitung von ML zuspitzt. Sobald sinnhafte Phänomene wie Denken oder Kommunikation als berechenbar vorstellbar werden oder sich gar auf rechentechnischer Basis hervorbringen lassen[42], ermöglicht dies nicht nur neuartige Perspektiven auf die technisierten Phänomene[43], sondern bringt ebenso praktische Probleme wie auch theoretisch anspruchsvolle Fragestellungen mit sich. Diese liegen darin begründet, dass – entsprechenden Suggestionen und Zuschreibungen zu trotz – lernfähige Rechenmaschinen nicht im Medium Sinn operieren und deren Outputs folglich missverstanden wären, nähmen Beobachter:innen dies an. Solche Missverständnisse könnten etwa eine Personalisierung von Maschinen zur Folge haben[44], die angesichts erwartbarer Erwartungsenttäuschungen wohl mit einer Vielzahl nicht-intendierter Nebenfolgen einherginge.[45]

Deshalb gehe ich im Anschluss an Weaver und Parisi von einer *epistemischen Lücke* zwischen adaptiver Rechentechnik und Sinnsystemen aus. Weil diese in unterschiedlichen Medien kognitiv operieren, ist auf ebendieser Ebene ihrer Operationen grundlegend von einer wechselseitigen Inkommensurabilität auszugehen. Die Beteiligung von ML an Kommunikation ist deshalb ein voraussetzungsvolles Anliegen, kann Sinn der Maschine[46] doch nur in entsprechend aufbereiteter (maschinenlesbarer) Formatierung ›vermittelt‹ werden. Und Gleiches gilt für die ›andere Richtung‹: nicht zufällig wird in »künstlicher Kommunikation« nicht die mathematische Lösung einer algorithmischen Berechnungssequenz beobachtet, sondern allein deren Resultat in ›übersetzter‹, d.h. sinnförmig vermittelter Gestalt.

Wenngleich die Outputs beider für Beobachter:innen informativ sein können – hierin besteht die interessante Parallele –, heißt dies weder, dass sie aufeinander reduzierbar wären, noch dass eine Möglichkeit der bruchlosen praktischen Integration beider unmittelbar in Aussicht stünde, etwa in Form einer Sprache, die

42 In dem Sinne, dass etwa die die Produktion von Outputs algorithmischer Modelle als Denken oder Kommunikationsbeteiligung beobachtet wird.

43 Die neuartige philosophische Reflexionen etwa über den Stellenwert induktiver, deduktiver und abduktiver Momente des Denkens angeregt hätten (vgl. Parisi 2018., S. 89f.).

44 Im Sinne der Zuschreibung von Personalität an Maschinen als eigenständiger soziologischer Sachverhalt, vgl. Harth 2014.

45 Spike Jonzes Film »Her« (Jonze 2013) stellt ein solches Szenario im Kontext der sich gewissermaßen nur einseitig entwickelnden Romanze zwischen Theodore Twombly und dem Sprachassistenzsystem Samantha in seinem tragischen Potenzial eindrücklich vor.

46 Genauer: den Beobachter:innen ihrer Outputs.

alle Eigenheiten beider vollständig zu berücksichtigen vermag. Diesseits des Phantasmas derartiger Symmetrie, die eine Einebnung der Unterschiede von menschlicher und maschineller Sozialität ermöglichen könnte, ließe sich indes schon an diesem Punkt fragen, mit welchen theoretischen Mitteln die geteilte Sozialität von Menschen und Maschinen adäquat beschrieben werden kann.[47] Als neue semantische Formen könnten diese begrifflichen Ressourcen jedenfalls solche augenscheinlich suboptimalen Beschreibungen wie die (soeben so genutzten) Ausdrücke »Übersetzung« oder »Vermittlung« zur Bezeichnung der Herstellung von pragmatischer Kompatibilität zwischen Menschen und Maschinen ersetzen.

In diesem Sinne möchte ich im Folgenden versuchen, zur weiteren soziologischen Erörterung des Untersuchungsgegenstandes den sozialen Charakter des Lernens in ML, d.h. soziologisch relevante Aspekte algorithmischen Modellierungsverfahren näher zu bestimmen. Diese Ausführungen sollen einerseits dem Anliegen der Annäherung an eine sozialtheoretische Konzeption von ML inkl. der Bestimmung seiner sozialen Funktion dienen. Andererseits soll es mit Hinblick auf die empirische Untersuchung der Herstellung und Anwendung algorithmischer Modellierungen dazu beitragen, für Eigenheiten und Fallstricke der praktischen Umsetzung von ML zu sensibilisieren.

Wie oben bereits erwähnt, ist der Vorgang einer algorithmischen Modellierung als Teil von ML mathematisch am besten verstanden als automatische »Operation des Findens einer Funktion« (Mackenzie 2017, S. 80, Übers. RG) zur Beschreibung eines Datensatzes mittels eines Algorithmus. Errechnet wird dabei eine mathematische Funktion, die infolge der Modellierung so exakt wie möglich die Zusammenhänge zwischen Variablen oder Wertmengen als Vektoren in einem n-dimensionalen Raum beschreibt (vgl. ebd.).[48] Die rekursive Annäherung – Rechenschritt für Rechenschritt – an eine möglichst präzise Beschreibung des Datensatzes erfolgt als

47 Angesichts dieser Disposition ist zu betonen, was Petter und Anton Törnberg (2018) als Warnung vor einer Versuchung für die (nicht nur soziologische) Technikforschung formulierten, dass nämlich im Zuge der Soziologisierung (bzw. geisteswissenschaftlichen Erforschung) von Technik sich unter der Hand eine Naturalisierung des Sozialen vollziehen könnte. Tendenzen dazu könnten etwa darin erkannt werden, dass im Anliegen etwa der Dekonstruktion eines verklärten »algorithmic sublime« (Ames 2018) manche sozial- und geisteswissenschaftlich fundierte Kritik an »KI« mitunter auf eine allzu selbstsichere (und womöglich gar ihrerseits sublim anmutende) Version der ›wahren‹ und/oder ›richtigen‹ gesellschaftlichen Ordnung baut. Hier stößt die Analyse auf die Verantwortung menschlicher Akteur:innen im Zusammenhang mit Formen von Computertechnik, über die in Anlehnung an David Gugerli (2018) gesagt werden kann, dass die Welt längst in ihnen angekommen zu sein scheint. Die Krux infolge dieses Umstandes besteht – praktisch betrachtet – eben darin, dass die Welt *in* dieser Technik (etwa in Form eines algorithmischen Modells) nie die gleiche ist wie jene, in der die Technik zum Einsatz kommt.

48 Die Anzahl der Dimensionen dieser Vektorenraums (*vector space*) hängt von der Anzahl der relevanten *features* im Datensatz ab.

iterative Optimierung (»[function] fitting«, vgl. Mackenzie 2017, S. 86 bzw. räumlich: »curve fitting«, vgl. Marcus 2018) der durch die Funktion beschriebene Kurve an die Datensatzstruktur. Der Maßstab für die ›Nähe‹ der Näherung an die Datenpunkte ist dabei eine »cost« bzw. »loss function«, die als Indikator für den Grad der Optimierung der Funktion dient und hinsichtlich der Anzahl an Fehlern (etwa in Klassifizierungen oder Prädiktionen) iterativ minimiert werden soll (vgl. Mackenzie 2017, S. 96).[49]

Die ›Wirklichkeitstreue‹ des errechneten Modells kann nicht modellintern, sondern nur im Wirklichkeitskontakt geklärt werden. Dazu können im Fall von analytischen Verfahren Tests mit neuen (Wirklichkeits-)Daten vorgenommen werden zur Einschätzung der Validität von ›gelernten‹ Klassifikationen. So kann ebenso untersucht werden, ob das Modell möglicherweise zu *overfitting* oder *underfitting* tendiert. Diese Unterscheidung dient der Beschreibung des Zusammenhanges von Modell, Datensatz und Wirklichkeit (vgl. Mackenzie 2017, S. 125). Ein »Overfit« bezeichnet eine übermäßig detaillierte Modellierung, die zwar die Eigenschaften des Datensatzes überaus präzise beschreibt, diese jedoch nicht repräsentativ für das eigentlich zu modellierende Phänomenen sind. Ein solcher Fall liegt etwa vor, wenn ein Bilderkennungsmodell (mit ›Universalanspruch‹ auf) alle Inputbilder von jungen Menschen mit dem Label »Schauspieler:in« versieht, weil sich im Datensatz überwiegend (auch derart gelabelte) Bilder von Schauspieler:innen befinden. Das Modell weist demnach – gemessen an der modellexternen Realität – einen zu hohen Grad an Spezifizität auf und bildet den Datensatz akkurat ab, ohne dass damit die anvisierte ebenso akkurate Abbildung des interessierenden Phänomens gelänge. Ein »Underfit« liegt entsprechend im gegenteiligen Fall vor, wenn die Kategorien eines Modells zu wenig spezifisch verfasst sind, um die »ground truth« des zugrundeliegenden Datensatzes zum Ausdruck zu bringen. Bezogen auf das eben gewählte Beispiel entspräche dies einer Disposition, in der sich mittels des Modells keine akkurate Zuordnung von Bildern junger Menschen erreichen lässt und diese stattdessen etwa als Bilder von alten Menschen oder etwa Bäumen klassifiziert werden.

Im Fall von generativen Anwendungen kann anhand der Outputs die Realität des Modells beurteilt werden. Dies mag etwa im Fall einer ›Befragung‹ von LLMs auf faktisch verfügbares Wissen schon leichter sein, als wenn man diese auf ihre ›Einstellungen‹ zu kontroversen Themen und mithin ihre normativen ›Vorurteile‹ (»Bias«) testet.[50] Noch schwieriger wird eine Beurteilung allerdings, wenn von einem Mo-

49 Die Kostenfunktion bezieht sich dabei allein auf die Beziehung von Modell und Datensatz. Sie lässt deshalb keine Aussage über die Generalisierbarkeit der Modellimplikationen auf realweltliche Zusammenhänge zu.

50 Es zeigt sich, dass ›Safeguarding‹-Strategien der verantwortlichen Unternehmen zunehmend darauf abzielen, solche Themen (manuell) zu ›sperren‹, sodass die Modelle auf heikle Fragen mit generischen, ausweichenden bzw. unverbindlich-diplomatischen Antworten reagieren. Dabei ist wichtig zu sehen, dass die Modelle nicht ›selbst‹ etwa über eine Unterscheidung von

dell ›kreative‹ Outputs erwartet werden – fotorealistische Genauigkeit von Bildern oder inhaltliche Vollständigkeit von faktischen Angaben in Texten mögen sich dann als deutlich leichter zu approximierende Kriterien erweisen als künstlerische oder poetische Qualität. Abgesehen vom speziellen Fall der Sprachmodelle müsste ohne hin von Anwendungsfall zu Anwendungsfall beurteilt werden, wie ein Validitätstest plausibel konstruiert werden kann.

Typen algorithmischen Lernens

Im Sinne einer typologischen Heuristik werden üblicherweise drei verschiedene Arten von ML-Ansätzen unterscheiden, in denen sich jeweils distinkte ›Lernverfahren‹ manifestieren: 1. überwachtes, 2. unüberwachtes (*[un-]supervised*) und 3. Verstärkungs- bzw. Belohnungslernen (*reinforcement learning*) (vgl. Mackenzie 2017, S. 80). Der Supervisionsaspekt besteht bei ersterem etwa durch Festlegung interessierende Korrelationen, indem eine explizite Klassifizierung der Trainingsdaten vorgenommen wird. Die Modellierung erfolgt also mit Daten, in denen die Korrelation von Input (Bild) und Output (Label) enthalten ist – wie etwa im Fall von ImageNet (vgl. Kap. 1, S. 11f.) wird das Modell im Hinblick auf die im Training zu lernenden Kategorien prädeterminiert.

Bei unüberwachten Verfahren erfolgt die algorithmische Mustererkennung im Sinne des automatischen Errechnens einer Funktion gewissermaßen frei und die ›gefundenen‹ Muster können im Anschluss an die vorgabenfreie Modellierung ggf. zu Kategorien erklärt werden. Beispiele für diesen Ansatz sind Verfahren des Clusterings und der Anomalieerkennung in Datensätzen oder solche, die deren ›Hauptkomponenten‹ (*PCA – Principal Component Analysis*) identifizieren. Als ein Beispiel für solche Verfahren kann Googles »Artificial Brain«-Modell dienen, das 2012 für Furore sorgte, als es mutmaßlich ›von sich aus‹ begann, auf YouTube ›Katzenvideos‹ zu clustern, d.h. diese zuverlässig von anderen Videos zu unterscheiden (vgl. Clark 2012). In diesem Fall konvergierten gewissermaßen menschliche Wahrnehmung und *computer vision*, d.h. statistische Analysen von Bildpunkt-Häufigkeitsverteilungen, sodass der Erfolg des ›freien Trainings‹ leicht erklärbar war. Generell gilt jedoch insbesondere für unüberwachte Verfahren, dass Training ohne »direct measure of success« (Mackenzie 2017, S. 80) in eine Disposition führt, in der schließlich menschliche Beobachter:innen interpretativ darüber entscheiden müssen, was genau das Modell eigentlich ›erkannt hat‹ – sofern die errechneten Muster sich überhaupt als verstehbar erweisen. Dennoch sind unüberwachte Verfahren schon aus einem pragmatischen Grund in vielen Anwendungsszenarien populär: Datensätze, die als *ground truth* für die Modelle dienen, liegen für häufig nicht in aufgearbeiteter (etwa klassifizierter) Form vor (vgl. Sudmann 2018a, S. 11) – oder

kognitivem (wirklichkeitsbezogen ›faktisch‹ und daher wissenschaftlich korrigierbar) und normativem Wissen (wertbezogen) verfügen, die modellintern differenzieren könnte (vgl. für diese Unterscheidung Luhmann 1991, S. 138ff.).

sind etwa aus Datenschutz- oder Kostengründen nicht für ML-Anwendungen verfügbar.

Reinforcement learning (RL) beruht dementgegen auf wesentlich anderen Grundprinzipien. Verfahren des RL nehmen in der Regel einen virtuellen Agenten an, der für ›Entscheidungen‹ positiv oder negativ sanktioniert wird und in diesem Zuge eine Strategie entwickelt.[51] Paradigmatisch für einen solchen Ansatz steht klassischerweise die ›Skinner-Box‹, in der mit Tieren (insbesondere Ratten) Experimente zu Mechanismen der Verhaltenskonditionierung durchgeführt werden. Insofern verweisen RL-Verfahren im Hinblick auf die implizierten Konzeptionen des Lernens auf den Behaviorismus in Ethologie und Psychologie seit Ivan P. Pavlov und Edward. L Thorndike (vgl. Sudmann 2018a, S. 19). RL spielt im Kontext dieser Studie keine maßgebliche Rolle, da es weder in den konkreten empirischen Fälle der Untersuchung noch in den beschriebenen Trends um ML der letzten gut zehn Jahre von wesentlicher Bedeutung ist. Eine Ausnahme bilden jedoch Spiele und vergleichbare Kontexte, die nicht auf die Modellierung eines sozialen Phänomenzusammenhangs abzielen, sondern auf die ›Ausbildung‹ eines ›möglichst intelligenten‹ isolierten Agenten abzielen, der schrittweise (Spiel-)Züge vornimmt (vgl. ebd., S. 11).[52] RL wird in dieser Studie auch deshalb kaum Erwähnung finden, weil mit der Agentenfokussierung des Ansatzes in behavioristischer Anlage tatsächlich wesentlich andere sozialtheoretische Implikationen einhergingen.[53] Dies gilt zumindest hinsichtlich der Sozialität (aus der Perspektive) des Agenten, ohne dass dadurch allerdings das unter der Beteiligung (von Bots) sich ergebende soziale Geschehen prädeterminiert wäre.[54]

Allen Formen von ML ist indes gemein, dass die konkrete Art und Weise des Lernens von einer mathematischen Funktion definiert wird, dem Lernalgorithmus. Darüber hinaus ist eine weitere mathematische Funktion das Ziel; jene, die das ›Gelernte‹ beschreibt, ob Häufigkeitsverteilungen oder Zugfolgen: das Modell. Von Verfahren des *reinforcement learning* abgesehen ist die (soziale) Funktion dieser (mathematischen) Funktion die Identifikation von statistischen Mustern. *Lernen* erscheint

51 RL wird aufgrund der maßgeblichen Konditionierungsinstanz mitunter auch zu »semi-überwachten Lernverfahren« (ebd.) gezählt.

52 Etwa im Videospielkonsolenklassiker Atari, vgl. Mnih et al. 2013.

53 Vgl. für das Argument, dass in RL in aller Regel eine sozialtheoretische Reduktion auf Überlegungen der Spieltheorie stattfindet: Roberge & Castelle 2021, S. 6 sowie ausführlicher, vergleichend und weiterführend zur ›Perspektive‹ von ML-Modellen ›auf die Welt‹: Castelle 2020.

54 Siehe Jonathan Harths Studie zu verschiedene Formen der Zuschreibung von Trivialität oder aber Personalität gegenüber Bots in Computerspielsituationen untersucht werden. Eine instruktive Points Harths betrifft dabei Feedbackeffekte: So lassen sich nicht nur (temporäre) Personalitätszuschreibungen an beobachten, sondern ebenso (Selbst-)Trivialisierungen der spielenden Menschen, vgl. Harth 2014, S. 307f.

hier synonym mit dem (errechnenden) Finden einer adäquaten mathematischen Funktion, die dies leistet (vgl. Mackenzie 2017, S. 82):

> We wish to know: in what sense does a machine learner learn? This question can […] be reframed: how do machine learners find functions? (ebd., S. 83)

Charakteristisch für das Training bzw. Lernen in ML ist daher ein Prozess des Findens bzw. des »fitting«[55], der Daten als Funktionen bzw. Funktionen in die Daten bringt (vgl. ebd., S. 86). Dieser Zusammenhang allein allerdings ist nicht nur für ML charakteristisch, sondern zeichnete schon Methoden statistischer Inferenz in der ersten Hälfte des 20. Jahrhunderts aus, etwa Verfahren logistischer Regression, die ebenfalls im ›dazu im Stande sind‹, Datenpunkte mittels einer Trennungslinie zu klassifizieren entsprechend der Maßgabe, dass diese den bestmöglichen »fit« gemessen an der Verteilung der Datenpunkte darstellen sollte (vgl. Mackenzie 2015, S. 433). Der wesentliche Unterschied zwischen ML und »classical statistics« besteht darin, dass ML nicht mehr den gleichen Beschränkungen hinsichtlich der Größenordnung der zu verarbeitenden Daten unterliegt: »the obstacle of the scale of data has shifted« (ebd., S. 434). Die aus den Vektoren zur Beschreibung unterschiedlicher Datenpunkte resultierenden Räume umfassten in konventionellen Verfahren etwa zehn Variablen, woraus sich entsprechend ein Vektorenraum mit der gleichen Anzahl von Dimensionen ergab, in dem ein Sample von mehreren Tausend Datenpunkte ›untergebracht‹ war. Dementgegen umfassen die Vektorenräume von heutigen ML-Verfahren oft mehrere Zehntausend Dimensionen und beinhalten Millionen, wenn nicht Milliarden von Datenpunkten. Der Untersuchung der Beziehungen einzelner Variablen zueinander entspricht daher heute die Analyse hochdimensionaler Muster, die potenziell eine Vielzahl an »features« enthalten kann (vgl. ebd.).

Die sich im ›Lernen‹ bzw. ›Training‹ vollziehende »Delegation« (Jaton 2021, S. 278) der Modellformulierung an Algorithmen lässt sich mithin als kognitiver Vorgang verstehen. Folgt man Dirk Baecker, der Kognition mit Heinz von Foerster als »Angelegenheit rekursiver Berechnung von Berechnungen« (Baecker 1995, S. 167) operativ geschlossener Systeme[56] versteht, zeigt sich, dass dies tatsächlich eine adäquate Beschreibung für das »function finding« (Mackenzie 2017) im Sinne einer algorithmischen Optimierung der Beschreibung eines Datensatzes darstellt (vgl. Baecker 1995, S. 166ff. sowie von Foerster 1993b/c). Ein soziologischer Begriff von Kognition wäre im Anschluss daran über den von Weaver herausgearbeiteten doppelten Begriff von Kommunikation bzw. Information zu plausibilisieren. So ließe sich der Sinnbezug von Kognition markieren, ohne dass die Kognition ihrerseits sinnförmig ablaufen müsste: »Cognition is a process that interprets information

55 Worauf auch das oben erwähnte Konzept des »Overfitting« (vgl. S. 65) bezogen ist.

56 Dirk Baecker betont in diesem Zusammenhang den engen Zusammenhang zwischen Kognition als fundamentlose »Bewegung im Netzwerk prinzipiell unendlicher Kontingenzen« 253) und dem Beobachtungsbegriff (vgl. Baecker 2013, S. 253).

within contexts that connect it with meaning« (Hayles 2017, S. 22), heißt es etwa bei Katherine Hayles.[57]

ML kann so soziologisch als *kognitive Technik* verstanden werden und es ist mithin gerade der rekursive Charakter des Modellierungsvorgangs, der sich gewissermaßen als Fluch und Segen zugleich erweist: Weil sich die algorithmische »Berechnung von Berechnungen« (Baecker 1995, S. 167) – das Modell – in aller Regel nicht präzise genetisch rekonstruieren lässt (vgl. Sudmann 2018a), verbietet sich auch ein exaktes, auf einzelne Rechenoperation hin aufgelöstes Verständnis dieses kognitiven Vorgangs. Ein solches wäre zweifellos von großem wissenschaftlichen wie praktischen Interesse und könnte einen wesentlichen Schritt in Richtung ›Explainable AI‹ (bzw. hier genauer: »Model Interpretability«, vgl. Lipton 2018) ermöglichen, lässt sich zum gegenwärtigen Zeitpunkt für ML jedoch (noch immer) nicht ausmachen. Weil mit dem opaken Charakter der technischen Vorgänge allerdings nicht zuletzt auch sozusagen ›epistemisches Potenzial‹ einhergeht[58], ist wohl davon auszugehen, dass auch alles im Sinne praktischer Nützlichkeit von ML mit *Intelligenz* Assoziierte gleichsam nur um den Preis von Intransparenz und mithin begrenzter Erklärbarkeit zu haben ist.[59] Inwiefern sich ML-basierte Technik in einem konkreten empirischen Szenario als praktisch nützlich erweist, hängt indes davon ab, ob sich deren Outputs[60] – die von einem algorithmischen Modell produzierten Datenanalysen (vgl. Dourish 2016) – als *Beobachtungen* sozial anschlussfähig zeigen. In anderen Worten: ML bewährt sich sozial durch seine kognitiven Fähigkeiten – seine Outputs werden als Beobachtungen beobachtet. Ausgehend von dieser Disposition soll die soziologische Annäherung an ML nun eine neue Richtung einschlagen, die dessen Status als Technik beleuchtet.

---

57 Die an gleicher Stelle zudem explizit auf Shannons und Weavers Arbeiten verweist. Mit dieser Definition grenzt sich Hayles zugleich von mechanistischen Beschreibungen geistiger Prozesse (›mind/brain is a computer‹) ab, da Information und Sinn explizit nicht gleichgesetzt werden (vgl. Hayles 2016, S. 786). Ihr Begriff von Interpretation (»process that interprets information«) zielt allein auf die Selektivität kognitiver Vorgänge ab (vgl. Hayles 2017, S. 25).

58 Man denke nur an gegenwärtige Versuche, interaktive Anwendungen wie Siri oder Alexa zu umfassenden ›personal assistants‹ weiterzuentwickeln, die auf (Sprach-)Kommando eine zunehmende Bandbreite an ›tasks‹ unmittelbar erledigen können.

59 Anders gesagt: »opacity – ›no direct measure of success‹ – is generative in machine learning« (Mackenzie 2017, S. 81). Eine ausführliche Auseinandersetzung mit den Begriffen Intelligenz und Lernen samt Überlegungen zum Zusammenhang mit Transparenz, Verständlichkeit und Trivialität folgt in Kap. 6.

60 Wobei schon die den Beobachtungen zugrundliegenden Funktionen als Beschreibungen des Datensatzes Beobachtungen darstellen: »Viewed as an enunciative function, machine learning makes statements through operations that treat functions as partial observers« (ebd., S. 81).

Zwischen Korrelation und Kausalität

Dazu erfolgt zunächst eine Bestandsaufnahme: ML wurde am Beispiel des Bilderkennungsalgorithmus von Krizhevsky et al. aus dem Jahre 2012 unter Verweis auf das *ImageNet*-Projekt eingeführt. Es wurde anhand dieses sowie anderer Beispiele in Grundzügen dargelegt, dass ML auf der algorithmischen Modellierung eines Datensatzes beruht. Der Maschine im Kompositum ML entspricht daher kein mechanisches Arrangement, sondern ein algorithmisch errechnetes Modell, das im Sinne einer Minimaldefinition nicht mehr als eine »compact, digested form of the information latent in the training data, in a form suitable for making decisions« (Rao 2021) ist. ›Die Maschine‹ erscheint daher in erster Linie als »conceptual device« (von Foerster 1984, S. 9).

Mittels eines geeigneten Inferenzalgorithmus können einem solchen Modell mittels geeigneter Inputs Outputs entnommen werden. Die Bezeichnung »Maschine« ist deshalb einleuchtend, weil das Modell schließlich – in Form von Software mit dem Inferenzalgorithmus verknüpft[61] – als Computerprogramm genutzt werden und dabei bestimmte Funktionen erfüllen kann. So konnten nach Abschluss der Verarbeitung des *ImageNet*-Datensatzes – bestehend aus klassifizierten (›gelabelten‹) Bildern – unter Einsatz des von Krizhevsky et al. entwickelten *Deep Convolutional Neural Network* auf zuverlässige Art neuen Inputs (Bildern) zum Datensatz passende Outputs (Labels) zugewiesen werden. Der zweite Teil des Kompositums ML – *Lernen* – bezieht sich mithin praktisch auf die Befähigung zur ›korrekten‹ Repräsentation der Inhalte des dem Modell zugrundeliegenden Datensatzes. Dies ist beobachterrelativ zu verstehen: So wie sich eine beobachter:innenunabhängige ›objektive‹ Feststellung des Sinns eines Datensatzes soziologisch kaum plausibilisieren lässt, gilt dies notwendigerweise auch für die Beurteilung des Lernerfolgs eines Modells nach Abschluss von dessen *Training*. ›Gelernt‹ wird in diesem Sinne qua algorithmischer Errechnung eine mathematische Funktion zur Beschreibung des Datensatzes.

Technik und Ungewissheit[62]

Zentral ist in diesem Zusammenhang die Bedeutung von Wahrscheinlichkeit. Die algorithmische Modellierung eines Datensatzes entspricht streng genommen einer (quantitativen) Datenanalyse, auf deren Basis in der Folge weitere Daten(-analysen)

61 Im Folgenden werde ich mich hinsichtlich ML mit der Bezeichnung »[algorithmisches] Modell«, sofern nicht explizit anders ausgewiesen, pragmatisch auf die Kombination aus Modell und Inferenzalgorithmus beziehen.

62 Teile des folgenden Arguments wurden bereits in Groß 2025 publiziert. Der Abschnitt stellt eine überarbeitete und erweiterte Fassung der dort entwickelten Überlegungen dar.

produziert werden können (vgl. Dourish 2016, S. 7). Die Analyse besteht in der Erfassung einer Verteilung von Datenpunkten als Beschreibung von deren Beziehungen, von der aus auf weitere Beschreibungen ›geschlossen‹ (im Sinne statistischer Inferenz) werden kann. Im Falle von *AlexNet* geht es um den Zusammenhang von Bildern als Rekombinationen von Pixeln und Wörtern, die zu deren Bezeichnung vorhanden sind.[63] Auf Basis der erfolgten Analyse kann *AlexNet* eine Angabe dazu errechnen, mit welcher Wahrscheinlichkeit ein Bild einer in *ImageNet* definierten Klasse angehört. Bei Festlegung eines Grenzwertes (etwa 90 % Wahrscheinlichkeit im Modell) kann diese Angabe als Tatsache präsentiert werden[64]: ›Katze‹, ›Hund‹ – oder aber auch ›nichts erkannt‹. Wobei dies schon unsauber formuliert wäre, schließlich *sieht*, geschweige denn *erkennt* das Modell nichts, zumindest nicht im konventionellen, zumeist auf Verstandesleistungen bezogenen Sinne. Stattdessen beruhen die Modelloutputs allein auf Wahrscheinlichkeitsangaben, die sich auf die Analyse der Häufigkeitsverteilung von Pixeln in einem Bild beziehen.

Dieser Aspekt gilt weiterhin auch dann, wenn auf der Ebene des Interfaces für Nutzer:innen einer Bildklassifizierungsanwendung statt einer Auflistung wahrscheinlicher Ergebnisse einer Analyse nur das höchstwahrscheinliche aufgelistet wird, sodass für Beobachter:innen der Eindruck entstehen kann, dieses sei eindeutig *erkannt* worden. Eine solche Reduktion vom Komplexität zur Herstellung von Eindeutigkeit wäre dann als »interface effect« (Galloway 2012), nicht aber als grundlegendes Charakteristikum von ML überhaupt zu verstehen. Dieser Aspekt verdeutlicht allerdings die Verstrickung von ML mit anderen Aspekten des Designs wie auch der praktischen Nutzung technischer Geräte. Allein Interfaces – die zur Nutzung von Computertechnik wie ML unabdingbar sind – erweisen sich in ihrer konkreten Gestalt zugleich als Ausdrücke sozialer Normen und technischer Datenverarbeitung, die ihrerseits Konventionen unterliegt (vgl. ebd.).

---

63 Sog. Labels, die als Klassen/Kategorien des Modells dienen.

64 Weil es sich hierbei um eine Design- bzw. Arrangementfrage handelt, ist in diesem Zusammenhang terminologische Präzision wichtig. So erscheint etwa die Rede von den ›Entscheidungen des Algorithmus‹ soziologisch problematisch, weil ein Algorithmus offensichtlich keine Entscheidung treffen kann, sofern man eine solche auf ihre logischen Bedingungen hin mitsamt ihrer reflexiven Implikationen betrachtet. Heinz von Foerster hat die notwendige Selektivität und mithin den willkürlichen bzw. kontingenten und gewissermaßen paradoxen Aspekt von Entscheidungen auf die folgende griffige Formel gebracht: »Nur *die* Fragen, die im Prinzip unentscheidbar sind, können wir entscheiden« (von Foerster 1993a. S. 72). Dementsprechend kann das ›Abverlangen‹ einer Entscheidung von einem Algorithmus nur zu einer Fehlermeldung führen, ist jede algorithmische Rechensequenz doch logisch determiniert, sodass es ›für einen Algorithmus‹ nie zu einer »Wahl innerhalb von Alternativen« (Luhmann 1997, S. 831) kommen kann. Dass solche semantischen Formeln mitunter dennoch als plausibel akzeptiert werden, hängt folglich an Interfacedesigns, die solche Eindrücke erzeugen können, vgl. ausführlicher dazu die im Haupttext unmittelbar anschließenden Überlegungen zu Interfaces.

Der probabilistische Charakter von ML kann aus verschiedenen Gründen unsichtbar gemacht werden: Bewerber:innen um einen Kredit bei einer Bank etwa nützt es wenig, zu erfahren, mit welcher Wahrscheinlichkeit sie vom algorithmischen Modell als nicht kreditwürdig klassifiziert worden sind, solange jedenfalls ihnen die Angabe nicht dabei helfen kann, die Entscheidung anzufechten, um letztlich einen Weg zu finden, den gewünschten Kredit doch zu erhalten. Dies beschreibt die Erlebensseite einer Form der sozialen Beteiligung von ML. Von der Handelnsseite aus ergibt sich ein ähnliches Bild. Bekommen etwa Polizist:innen im Streifendienst mittels ML-basierter Prädiktionssoftware raumbezogene Angaben zur Wahrscheinlichkeit von Straftaten angezeigt, ist für die Folgen dieser Mitteilung – Priorisierung bestimmter Gegenden für Streifenfahrten – unter Umständen nachrangig, ob auf dem Display im Streifenfahrzeug eine eindeutige Angabe oder diese in Kombination mit einer weiteren Information zum modellinternen Konfidenzintervall der Prädiktion erscheint.[65] Ist die wahrscheinlichste (zumeist ›oberste‹) Option nämlich in die Regel ohnehin diejenige, auf die die Wahl fällt, hat eine eingeblendete Wahrscheinlichkeitsangabe womöglich gar keine praktische Relevanz.

Eine solches Anwendungsszenario erscheint gleichwohl dazu prädestiniert, selbsterfüllende Prophezeiungen und zu produzieren: Größere Polizeipräsenz in bestimmten Gegenden kann zu mehr festgestellten Straftaten führen, wobei der Effekt der Mitteilung unabhängig von ihrem Wahrheitscharakter maßgeblich zum Eintreten der Vorhersage beigetragen hätte.[66] Ein solches Szenario hängt einerseits an der generellen Konzeption einer Anwendung, etwa des »Predictive Policing«, andererseits allerdings auch maßgeblich am Design des genutzten Interface, das reguliert, in welcher Form Ergebnisse des probabilistischen Modells für Nutzer:innen präsentiert werden.

Sinn und Funktion von Technik

Wie lässt sich eine solche Disposition als Technik beschreiben? Die vorangehenden Überlegungen deuten darauf hin, dass der stochastische Charakter algorithmischer Modellierungen und mithin ihrer Outputs für Komplikationen sorgen könnte, wollte man ML als ›klassische‹ Technik fassen. Die beschriebene probabilistische Operationslogik scheint in jedem Falle kaum eine Identifikation mit dem ›mechanistischen Paradigma‹ von Technik zuzulassen, innerhalb dessen Technisierung kausa-

65 Die Modellierung der Wahrscheinlichkeit einer Straftat zu einer bestimmten Zeit an einem bestimmten Ort kann von Verhalten, Personen, Netzwerken und Orten ausgehen vgl. Kaufmann, Egbert & Leese 2019.

66 Weil zugleich in anderen Gegenden eine geringere Polizeipräsenz zu mehr tatsächlichen und zugleich weniger festgestellten Straftaten führen kann.

ler Fixierung gleichkommt. Dies gilt jedenfalls für die Beziehung von Inputs und Outputs in ML-Anwendungen. Bevor ich diesen Umstand technikheoretisch zu erörtern versuche, möchte ich zunächst anhand einer Reihe empirischer Beispiele veranschaulichen, worauf diese Überlegung abzielt.

Die beschriebenen, mit dem praktischen Einsatz von ML einhergehenden Ungewissheiten, etwa in der Polizeiarbeit, suggerieren dementgegen vielmehr, dass eine solche Identifikation zu folgenschweren Missverständnissen führen kann. Den stochastischen Charakter von Modelloutputs zu verkennen, könnte etwa die Verwechslung einer Prädiktion mit einer gewiss eintretenden Zukunft zur Konsequenz haben. Über eine Vielzahl denkbarer Anwendungsszenarien hinweg scheint der Einsatz von ML mit hohen kognitiven Anforderungen an die adäquate Beurteilung der Bedeutung von Modelloutputs einherzugehen. Dies heißt indes nicht, dass auf die Unwägbarkeiten von Modelloutputs in entsprechenden technischen Anwendungen unbedingt qua Interface- oder Softwaredesign derart hingewiesen würde, dass der fragwürdige epistemische Status der Outputs klar vermittelt wäre. Gleichwohl kann ML in eine Anwendungsform gebracht werden, die epistemische Gewissheit suggeriert. Mit einer Reduktion des Auswahlbereichs qua Design geht allerdings die Invisibilisierung von Alternativen einher, die vom Modell ebenso errechnet wurden. Dies kann in der jeweiligen Anwendung wiederum zu unbeabsichtigten Nebenfolgen führen.

Besonders augenscheinlich wird die Relevanz von Interfacefragen in Vorhersageszenarien, etwa wenn Kreditbanken aufgrund automatischer Antragsevaluationsprogramme potenzielle Neukund:innen ablehnen, die bei persönlicher Prüfung durch Bankmitarbeiter:innen vermutlich akzeptiert worden wären. Sofern die zuständigen Mitarbeiter:innen für die manuelle Bestätigung des errechneten Vorschlags keine Informationen etwa zum Konfidenzintervall der Modellrechnung, geschweige denn eine explizite Begründung erhalten, gibt es für sie keine echte Möglichkeit der Evaluation der Berechnung, sondern allein die einer manuellen Gegenprüfung bei Fehlerverdacht.[67]

Oder aber es ergeben sich unübersichtliche – und nicht nur rechtlich weitaus problematischere – Kausalitätsbeziehungen, wenn Schwarze Menschen in den USA zu längeren Haftstrafen verurteilt werden aufgrund des Einflusses von auf

67 Dass sich hinter der Intransparenz von im Finanzbereich verwendeten Modellen mitunter aus der ›analogen Welt‹ bekannte Mechanismen zu verbergen scheinen, zeigt beispielsweise ein Fall um die Beantragung einer Apple-Kreditkarte aus dem Jahre 2019. Einem seit Jahrzehnten verheirateten Mann wurde aus unerfindlichen Gründen automatisch ein um ein Vielfaches höherer Kreditrahmen eingeräumt als seiner Ehefrau, obwohl beide seit 20 Jahren gemeinsame Steuererklärungen einreichen und sie zudem über einen besseren Credit Score verfüge, vgl. Vincent 2019.

einer überschaubaren Anzahl von Metriken beruhenden Prädiktionen von Rückfallwahrscheinlichkeiten auf die Urteile der zuständigen Richter:innen. Eine solche Form von »racial bias« wurde etwa für das infolge kritischer journalistischer Berichterstattung mittlerweile berüchtigte COMPAS-System festgestellt: Schwarze Menschen wurden verglichen mit weißen Menschen doppelt so häufig fälschlicherweise als zukünftige Straftäter:innen eingestuft (vgl. Angwin & Larson 2016).[68] Wohl kann dabei das Verhängen längerer Haftstrafen wiederum einen Einfluss auf die tatsächliche Rückfallquote haben. Sofern etwa lange Haftstrafen den Reintegrationsprozess der Inhaftierten im Anschluss an die abgesessene Zeit erschweren und deshalb die Wahrscheinlichkeit von Straffälligkeit erhöhen, hätte die Prädiktion erneut einen – im kybernetischen Sinne – positiven Feedbackeffekt erzeugt, d.h. zu ihrem Eintreten maßgeblich selbst beigetragen.

Die erwähnten Beispiele zeugen von – gemessen an den Absichten hinter ihrem Einsatz – fehlgeleiten ML-Anwendungen. Sie legen praktische Missverständnisse über die Funktionsweise von ML offen. Diese Missverständnisse dokumentieren nicht nur womöglich maßgebliche Aspekte des gesellschaftlichen Charakters von ML, sondern erweisen sich ebenso als Ansatzpunkte für soziologische Theoriebildung zu ML. Ich hatte zu Beginn dieses Abschnitts bereits angedeutet, dass ML sich kaum mit mechanistischen Vorstellungen von Technik identifizieren lässt, weil angesichts der Schwerverständlichkeit algorithmischer Operationen im Zuge sowohl der Modell- wie auch Outputerzeugung in ML typische Aspekte kausaler Fixierung nicht unmittelbar auszumachen sind.

Daran anknüpfend möchte ich nun mittels einer systemtheoretisch-funktionalistischen Argumentation die Frage nach einer adäquater Konzeption von ML als Technik vertiefen. Als Ausgangspunkt dieses Arguments soll das systemtheoretische Verständnis von Technik als »funktionierende Simplifikation« (Luhmann 1997, S. 524) dienen. In diesem Verständnis ist der Sinn von Technik »ihr Funktionieren« (Halfmann 2003, S. 136). »Was funktioniert, das funktioniert« (Luhmann 1997, S. 530) und bewegt sich darüber hinaus – anders als Kommunikation[69] – jenseits von Konsens und Dissens, denn: »Was sich bewährt, das hat sich bewährt. Darüber braucht man kein Einverständnis mehr zu erzielen« (ebd.). Dies gelte für klassische Werkzeuge wie den Hammer und gleichermaßen für »Informationsverarbeitungstechnik« (ebd., S. 525), sofern bei dieser distinkte »Konditionalgrenzen« (ebd.) im Sinne eines Programms vorausgesetzt werden können.

Technik erlaubt »gelingende Reduktion von Komplexität« (ebd.). Die Kriterien des Gelingens sind bei Beobachter:innen von Technik zu finden, d.h. in deren Erwartungen darüber, was ihr Einsatz bewirken werde. Führt die Anwendung von Technik

68 Der Fall erregte einiges öffentliches Aufsehen, in dessen Folge Pro Publica Kritik am methodischen Vorgehen der Studie erfuhr, vgl. Bechtel & Lowenkamp 2016.

69 Technik liege deshalb »gewissermaßen orthogonal zur Logik sozialer Systeme« (ebd., S. 524).

zuverlässig zu den erwarteten Resultaten, kann ihr die soziale Funktion der »Substitution von Konsens oder Dissens« (Halfmann 2003, S. 196) zugeschrieben werden, weil sie »die Beachtung mitlaufender Sinnbezüge (etwa über ihre Herstellungs- und Funktionsweise) entbehrlich macht« (ebd.). Dies bedeutet freilich weder, dass Technik in ihrer praktischen Anwendung tatsächlich zu beabsichtigten Ergebnissen führt, noch dass sich Komplexität durch sie auflösen lässt. Sozialwissenschaftliche Technikforschung hat gezeigt, dass genau das Gegenteil als Regelfall gelten kann und sich zumeist eine maßgebliche Divergenz zwischen (expliziten wie impliziten) im Design verankerten Suggestionen und tatsächlicher praktischer Techniknutzung auftut (vgl. Bijker/Law 1992). Nichtsdestotrotz bleibt der Aspekt der Erwartungsstabilität ein robustes Kriterium zur Charakterisierung von Technik. Jede Anwendung von Technik beruht auf der Erwartung, dass die Anwendung zu einem gewünschten Resultat führen wird. Dies ändert sich auch dann nicht, wenn sich das erwartete Resultat ändert. Diese kompakte funktionalistische Bestimmung deckt sich mit alltäglichen praktischen Erfahrungen, denn wenn Technik nicht mehr wie erwartet funktioniert und nicht die erwarteten Ergebnisse zu liefern imstande ist, erscheint sie ihren Nutzer:innen in aller Regel als kaputt.[70]

Ein solches soziologisches Verständnis von Technik bezieht sich nicht etwa auf deren ›Wesen‹, sondern findet die maßgebliche Referenz für Technik in der sozialen Wirklichkeit, in der sie zum Einsatz kommt. ›Dort‹ wird etwas als Technik bezeichnet, wenn sich in praktischer Nutzung zeigt, dass gleiche Inputs zuverlässig zu gleichen Resultaten führen und dies folglich auch mit Blick auf die zukünftige Nutzung erwartet werden kann. Technik erlaubt daher den Aufbau robuster Erwartungen, was die Ergebnisse ihres Einsatzes, d.h. ihr Funktionieren angeht.[71] Ein Bruch mit etablierten Erwartungen deutet in der Regel auf eine technische Fehlfunktion hin. Diese funktionalistische Konzeptualisierung legt nahe, dass die praktische Nutzung einer Technologie auf der Idealisierung der Erwartbarkeit ihres Funktionierens (auf

70 Damit soll keinesfalls das ästhetische oder epistemische Potenzial von technische Fehlfunktionen innerhalb experimenteller Praktiken bestritten werden (vgl. Rheinberger 2019), man denke mit Blick auf Kunst etwa an explizit technischen Fehlfunktionen verschriebene Genres (z.B. Glitch Art) oder an die Vielzahl von Geschichten gewissermaßen zufälliger, aber umso folgenreicher wissenschaftlicher Entdeckungen wie die von Dynamit durch Alfred Nobel oder von Penicillin durch Alexander Fleming. Robert K. Mertons und Elinor Barbers Studie zu »Serendipity« (vgl. Merton & Barber 2003) versammelt zahlreiche solche Geschichten glücklicher Zufälle, für deren Beschreibung sich diese semantische Formel gefestigt hat. Daran anknüpfen lässt sich wohl festhalten, dass es glückliche Umstände sind, die aus Unfällen ergiebige Einfälle stimulierende Zufälle machen.

71 Das bedeutet jedoch nicht, dass die Praxis selbst aufgrund ihrer technologischen Elemente stabiler oder robuster ist, wie es ›technosolutionistische‹ (vgl. Morozov 2013, Nachtwey & Seidl 2017) Weltanschauungen behaupten mögen.

unbestimmte Zeit) beruht. Dies gilt auch für den Fall technischen Störungen, jedenfalls dann, wenn sich im Zuge von Troubleshooting und Reparaturen Versuche der ›Wiederinstandsetzung‹ von Technik zeigen, die freilich auf der Annahme der prinzipiellen Möglichkeit einer Rückkehr zum gewünschten technischen ›Normalzustand‹ beruhen.

Zahlreiche ML-Anwendungsszenarien verkomplizieren m.E. ein solches Verständnis, wenn bei den Beobachter:innen der Maschine etwa gar keine verbindlichen Erwartungen darüber vorhanden sind, was deren Outputs im Normalzustand auszuzeichnen habe, weshalb es folglich schwierig sein kann, überhaupt zu beurteilen, ob die Maschine funktioniert oder aber kaputt ist. Outputs, die in »konventioneller« Software auf Fehlfunktionen oder Bugs hindeuten, könnten in ML-Anwendungen auf bislang unentdeckte/unverstandene Muster des Datensatzes verweisen. Mit dieser latenten Ungewissheit geht ebenso einher, dass sich die Behebung von durch Beobachter:innen als solche identifizierte Bugs einstweilen als diffizile, wenn nicht unmögliche Aufgabe darstellen kann. Was in ML ein Feature und was hingegen ein Bug ist, erweist sich Beobachter:innen nicht unbedingt als klar feststellbar. Deshalb meine ich, dass für Beobachter:innen – im Unterschied zu anderen Formen von Rechentechnik – das Auftreten von Unerwartetem nachgerade zu einem typischen Kennzeichen praktischer Einsätze von ML werden kann. Dass in diesen folglich also im Normalfall das Unerwartete zu erwarten sei, lässt fraglich erscheinen, ob ML als Technik im eben vorgestellten Sinne angemessen beschrieben ist.

Der Einsatz von ML führt häufig in eine praktische Disposition, in der stochastische Vorhersage von Beobachter:innen leicht mit kausaler Fixierung verwechselt werden kann, etwa aufgrund der Suggestionskraft des genutzten Interface. Auch wenn dieses derartige Eindrücke vermitteln kann: ML-basierte *Computer Vision* ›versteht‹ nicht, was sie ›sieht‹ und ›erkennt‹ nichts in Bildern, sondern errechnet Wahrscheinlichkeitswerte für die Übereinstimmung von Pixelverteilungen mit Kategorien, die als Klassen definiert sind und mit bestimmten Pixelmustern korrelieren. Auch wenn CV-Verfahren rein technisch betrachtet nahezu niemals ›falsch‹ liegen – i.S.v. ›sich verrechnet haben‹ –, kann gleichsam kaum angenommen werden, dass jedes relevante Bild mit Gewissheit korrekt identifiziert wird. Dies liegt daran, dass die rechentechnische Bilderkennung in Bezug auf ihre realweltliche Referenz nicht determiniert ist, sondern auf Basis einer stochastisch modellierten »ground truth« (Jaton 2024) Vorhersagen liefert, was für Beobachter:innen der Outputs freilich für lose Enden im Verständnis derselben sorgt. Die Beziehung zwischen Input (Bild) und Output (Kategorie) beruht auf nicht mehr und nicht weniger als statistischen Korrelationen zwischen als ähnlich identifizierten Inputs und mit diesen korrespondieren Outputs. Es handelt sich um Korrelationen, die *potenziell* realweltliche Kausalbeziehungen abbilden. Nichtsdestotrotz wird ML häufig verwendet, als könne kausale Determiniertheit in der Modellierung interes-

sierender Zusammenhängen vorausgesetzt werden. Eine funktional angemessene technologische Grundlage für eine Anwendung der Bildverarbeitung *könnte* zwar, *muss* jedoch nicht zwingend gegeben sein, da Beobachter:innen der Maschine zwar selbst Struktur und Inhalt des modellierten Datensatzes prinzipiell bekannt sein können, es jedoch selbst dann unmöglich bleibt, genau zu verstehen, wie diese im algorithmischen Modell repräsentiert werden.

Die wesentlichen praktischen Implikationen der konstitutiven Intransparenz algorithmischer Modellierungen lassen sich am einfachen Beispiel eines konventionellen Taschenrechners veranschaulichen. Über diesen kann angenommen werden, dass er vollständig logisch determiniert ist, d.h., auf der Ebene seines »subface« (Nake) »fest codiert« ist. Eine bestimmte Eingabe wurde so programmiert, dass sie eine bestimmte Ausgabe produziert, und wird dies daher auch in Zukunft tun, solange die Maschine funktionsfähig ist. Ein LLM hingegen wird bei einer Rechenanfrage angesichts seines »subface« ein Ergebnis generieren, das statistisch plausibel ist, d.h. dessen Position (als Datum) im Vektorraum des Modells ›in den Nähe‹ als ähnlich identifizierter (errechneter) Trainingsdaten. Ein LLM als Taschenrechner zu verwenden ist daher nur *potenziell* eine gute Idee.

Inwiefern lässt sich angesichts dieser Überlegungen ML als funktionierende Simplifikation verstehen? Auf Basis der vorangehenden Überlegungen drängt sich mir der Eindruck auf, dass ML nur partiell als Technik im eben angeführten Sinne verstanden werden kann. Einerseits scheint mir der Sinn von ML fraglos ebenso in seinem Funktionieren zu bestehen. In dieser Hinsicht erweist sich ML zweifellos als vergleichbar zu anderen technischen Formen inkl. digitaler Computer, und eine Vielzahl der gesellschaftlichen Anwendungsszenarien unterstützen diese Vermutung gewissermaßen als empirische Belege. Gewisse Zweifel hege ich allerdings mit Blick auf die Bestimmung der sozialen Funktion von ML. Von Technik, die Konsens oder Dissens substituieren und mithin Kommunikation redundant zu machen vermag, scheint mir ML in mindestens drei Hinsichten abzuweichen.

Erstens wäre in Zusammenfassung der bisherigen Ausführungen festzuhalten, dass die berichteten Kontroversen wie auch die Vielzahl von Skandalen angesichts unbeabsichtigter Nebenfolgen des Einsatzes von ML eine Basis für das Argument zu liefern scheinen, dass ML in mancher wesentlicher Hinsicht nicht hinreichend verstanden ist, dies allerdings vor allem auf seine intransparente und schwerverständliche Operationslogik verweist. Die operative Unzuverlässigkeit kann dabei sowohl als Ursache eines mangelnden Verständnisses der Operationslogik von ML verstanden werden wie auch als dessen Folge. Unabhängig von der Erklärung dieses Umstandes kann bereits die Beobachtung, dass ML-Anwendungen häufig keine zuverlässige Technisierung zu erlauben scheinen, als relevantes Indiz dafür verstanden werden, dass die soziale Funktion von ML gerade in der »Substitution von Konsens oder Dissens« (Halfmann/Luhmann) bestehen. Dementgegen scheint mir nachgerade das Gegenteil plausibel: statt Kommunikation hinfällig zu machen, sorgt ML

aufgrund seiner operativen Unzuverlässigkeit und Befremdlichkeit womöglich vielmehr für zusätzlichen Kommunikationsaufwand.

Zweitens lässt die bisherige Auseinandersetzung mit der stochastischen Operationslogik algorithmischer Modellierungen fragwürdig erscheinen, in welchen Zusammenhängen ML derart eingesetzt werden kann, dass Bedingungen von Erwartungsstabilität entstehen, wie sie für Technik typisch sind. Wiederkehrende Berichte davon, dass noch immer kaum professionellen Einsatzszenarien für ML gefunden sein – oft in Verbindung mit der Behauptung, KI/ML befinde sich (deshalb) im Stadium einer Blase kurz vor ihrem Zerplatzen dar (vgl. etwa Marcus 2023) – unterstreichen diese Vermutung. Es könnte sich deshalb mit Blick auf das »alien subject of AI« (Parisi 2019) womöglich zeigen, dass mein mit Rekurs auf Parisi und Weaver entwickeltes Argument der konstitutiven epistemischen Lücke in ML dessen Eignung als Technik verkompliziert, wenn nicht gänzlich ungeeignet erscheinen lässt.

Und drittens müsste insbesondere mit Blick auf sog. generative ML-Verfahren wohl gefragt werden, wie die soziale Funktion von deren Anwendungen bestimmt werden solle, wird ihnen doch nicht selten die Suggestion zugeschrieben, (›menschengleich‹) selbst Kommunikation zu stimulieren. Es wäre hier etwa kritisch zu vermuten, dass der generative Charakter rechentechnischer Kognition, wie sie sich in algorithmischen Modellierungen manifestiert, sich womöglich als ›zu generativ‹ für technische Zwecke erweisen könnte. Offensichtlich will und kann ich an dieser Stelle in all diesen Fragen keine Stellung beziehen. Mir scheint es angesichts der vorangehenden der techniktheoretischen Betrachtung von ML jedenfalls angebracht und für das sozialtheoretische Anliegen dieser Studie sogar überaus geeignet zu sein, die Frage nach der sozialen Funktion von ML ins Zentrum dieser Untersuchung zu stellen. Diese soll nicht nur die weitere theoretische Erörterung bestimmen, sondern ebenso die empirische Untersuchung anleiten.

Generative Stochastik

Instruktive Hinweise auf die soziale Funktion von ML sollten sich in besonders prägnanter Form gerade in solchen Anwendungsszenarien zeigen, in denen ML nicht im Sinne ›konventioneller‹ Technik eingesetzt wird bzw. wo ML die Technisierung von kognitiven Funktionen ermöglicht, die vorher kaum technisierbar schienen. Der nun folgende Abschnitt widmet sich sog. *generativen* Anwendungen von ML, bei denen das technische Funktionsprinzip gegenüber analytischen Verfahren (wie etwa *AlexNet*) gewissermaßen umgekehrt wird, wobei die analytische Komponente als Fundament der generativen Funktionen jedoch erhalten bleibt. Statt etwa allein Klassifikationen für Inputs zu leisten, sind solche Verfahren zur Produktion von Outputs imstande, die den Inputs (im Datensatz) ähneln und in der Folge wiederum analysiert werden, wobei die Wahrscheinlichkeit dafür berechnet wird, dass der

vom Modell erzeugte Output dem Trainingsdatensatz angehört (vgl. Goodfellow et al. 2016).

Ein weiterer Blick auf die vieldiskutierten LLMs vermag veranschaulichen, was mit der Anwendung einer solchen Rechenarchitektur einhergeht. Diese sind zur automatischen Textvorhersagefunktion fähig, die gemeinhin als »Autocomplete« bekannt sind. Fungiert ein hinreichend großes Textkorpus als *ground truth* eines LLM, kann näherungsweise für einen beliebigen Satz errechnet werden, was der nächste geschriebene Buchstabe oder das nächste Wort sein, oder aber auch, was der gesamte nächste Satz, der nächste Absatz oder gar das folgende Kapitel möglicherweise beinhalten wird. Grundlage der Berechnung dieser Prädiktion sind dabei sehr viele andere Texte. Dass auf »Ich« nicht selten »bin« folgt, mag trivial anmuten. Wie dieser aktuelle Absatz jedoch enden wird, ist im Moment des Tippens dieser Satzes selbst für den Verfasser dieses Textes nicht verlässlich bestimmbar. Diese introspektive Beobachtung kann wohl allemal als Indiz dafür hinhalten, dass es sich beim Fortschreiben eines Textstücks häufig um keine triviale Aufgabe handelt.

Zur Charakterisierung von LLMs mit Blick auf Funktionsweise und Output hat indes unlängst die Formel »spicy autocomplete« (Solomon 2023a) Verbreitung gefunden. Mit dieser beschreibt Mike Solomon infolge eigener Experimente, dass es sich entgegen manchen gegenteiligen Verlautbarungen – und eigenen ersten oberflächlichen Eindrücken – bei ChatGPT weniger um eine universell einsetzbare Wunderwaffe als vielmehr um eine ›aufgepeppte‹ Variante von Software zur automatischen Vervollständigung von Texten handelt, wie sie im Grunde schon längst bekannt ist (vgl. Solomon 2023b). Das zusätzliche Attribut »Spicy« leuchtet dann zur Kennzeichnung solcher maschinellen Texterzeugnisse ein, deren Verlauf nicht offensichtlich und folglich nicht ohne weiteres antizipierbar anmutet, weshalb die Texte ›originell‹ wirken können. Im Sinne der oben eingeführten Terminologie lässt sich »spicy« als Chiffre für ein nicht-triviales Verhältnis zwischen Input und Output deuten. Beobachter:innen der Outputs von ChatGPT und anderen LLMs dürften jedenfalls kaum umhinkommen, ein strukturierendes Prinzip zur Beschreibung des Vervollständigungsvorgangs zu unterstellen, seien es die internen Zustände der nicht-trivialen Maschine von Foersters (vgl. 1993b) oder das Bewusstsein einer *AGI* (*artificial general intelligence*). Die Einigkeit darüber, *dass* es sich um nicht-triviale Vorgänge handelt, die zu den Outputs von LLMs führen, heißt indes auch, dass sie wohl kaum als zufällig beobachtet werden. *Automatische Vervollständigung* meint in diesem Sinne zwangsläufig auch *automatische Ordnungsbildung*.

Für LLMs gilt dabei ›intern‹, dass allein statistisch plausible Ergebnisse ausgegeben werden. Diese können, müssen Beobachter:innen allerdings keinesfalls stimmig erscheinen. Die Frage nach der Stimmigkeit eines Outputs kann sich auf verschiedene Hinsichten beziehen. Sie kann etwa Syntax, Semantik, Ästhetik oder andere Hinsichten der Beurteilung von Texten betreffen, wobei das, was explizit ›in den Zeilen‹ steht (Syntax), verständlicherweise leichter statistisch zu verarbeiten ist

als das, was mitunter nur ›zwischen den Zeilen« zu finden wäre (Semantik). Alternativ könnte man diese Differenz als jene von Form und Inhalt reformulieren (vgl. Lopez 2023): Ein LLM-generiertes Textstück wird aufgrund seiner formalen Ähnlichkeit zu einer Vielzahl von Referenzobjekten in den Trainingsdaten vermutlich selten völlig aus dem Raster fallen und mag auf den ersten Eindruck hin häufig inhaltliche Plausibilität suggerieren, wenngleich immerzu erst eine exakte Prüfung Gewissheit verschaffen kann. »Dumme Bedeutung« nennt der Literatur- und Medienwissenschaftler Hannes Bajohr diesen Oberflächeneffekt von LLM-Outputs (vgl. Bajohr 2022). Bedeutsam auf den ersten, dumm auf den zweiten Blick – so könnten vermutlich zahlreiche Lektüreerfahrungen im Kontakt mit »artifizieller Semantik« (ebd.) ›aus der Feder‹ von LLMs beschrieben werden.

Derartige Erfahrungen implizieren einerseits letztlich eine Abwertung der Maschine auf Seiten der Text-Beobachter:innen, deren Enttäuschung wohl vor allem daher rührt, dass sie LLMs vorab überschätzt haben. Der sich in unrealistischen Erwartungen abzeichnenden Überhöhung der Maschine folgt im Zuge der Enttäuschung schließlich gewissermaßen ihre (Re-)Trivialisierung, zumindest hinsichtlich der expliziten Sinnbezüge ihrer Operationen (es gibt keine). Andererseits verweisen solche Begegnungen mit LLMs nicht nur auf Neues und Unverständliches. Insbesondere der generische Charakter von deren Outputs führt eindrücklich die Implikationen dessen vor, was das kulturelle Paradigma von »content generation« (Bull 2023) genannt werden könnte.[72] Dessen Emergenz beruht– und dies soll hier die wesentliche Pointe sein – auf einem epistemischen Bruch der einem mehrstufigen Prozess entspringt, zunächst der Verdatung von sinnförmig verfassten Phänomenen, der anschließenden algorithmischen Modellierung entsprechender Datensätze und schließlich der ›Rückübersetzung‹ in sinnhaft beobachtbare Outputs. Jede dieser Stufen birgt mit Blick auf den ursprünglichen Sinn transformatives Potenzial. So schreibt Bernard Rieder über den *Bayes Classifier* gewissermaßen stellvertretend für in ML zur Anwendung kommende Algorithmen:

> The Bayes classifier is thus neither a static recipe for decision-making nor a theory engaged in ontological attribution; it is a method for making data signify in relation to a particular desire to distinguish; it is a device for the automated production of interested readings of empirical reality. (Rieder 2017, S. 110)

72 Für dieses kann wohl als (proto-)typisch gelten, dass ›content creators‹ in Social Media generische Outputs produzieren, für deren Beurteilung in erster Linie das aufmerksamkeitsökonomische Kriterium der *Viralität* (höchstmögliche Anschlussfähigkeit) herangezogen wird. Dabei kann vermutlich als nachrangig angesehen werden, ob es sich bei den ›creators‹ um Menschen oder Maschinen handelt, auch wenn (gegenwärtig noch) die verbindliche Norm zu bestehen scheint, dass ›creators‹ als Personen adressierbar sein sollten.

Der »automatisierten Produktion interessierter Lesarten empirischer Wirklichkeit« fähige Technik vermag dazu beitragen, dass Daten (von Beobachter:innen) Bedeutung verliehen und ihnen mithin Sinn zugesprochen wird. Dies beginnt bereits auf der Ebene von Digitalisierung i.S.v. Verdatung: »Datafication […] translates fundamental assumptions about the application domain into data structures and reifies them« (ebd., S. 109). Derartige technikvermittelte Reifizierungen von Sinn in Datenform sind nicht mittels erklärender Rekonstruktion von Sinnzusammenhängen einholbar. Dies gilt für jede Ebene der sich vollziehenden epistemischen Transformation, die sich aus dem Zusammenspiel von digitalisierender Verdatung, algorithmischer Modellierung und statistischer Inferenz zusammensetzt, aus dem sich die rechentechnische Operationslogik von ML ergibt.

Wahrscheinlichkeit reflexiv

Zu dem beschriebenen konstitutiven Spannungsverhältnis zwischen Wahrscheinlichkeitsangabe über das Eintreten eines Ereignisses und dem tatsächlichen Ereignis gesellt sich in soziologischer Betrachtung unter Beachtung der realweltlichen Bedingungen sozialer Wirklichkeit eine weitere Komplikation. Diese besteht darin, dass keine prinzipielle Gewissheit darüber bestehen kann, wie faktisch stichhaltig die errechnete Wahrscheinlichkeitsangabe ist, da sie auf Basis eines Modells des Wirklichkeitsphänomens errechnet wurde, dessen Beschreibungskraft mit Blick auf das eigentlich interessierenden Phänomen nicht als garantiert angenommen werden kann. Es gibt keine Gewähr dafür, dass die errechnete Wahrscheinlichkeit den realen Zusammenhang adäquat erfasst – das Modell könnte derart konstruiert sein, dass es Rückschlüsse über die Wirklichkeit nicht zulässt. Dieser potenzielle Fehler – die Abweichung des Modells von der Wirklichkeit – wird auch als »Modellrisiko« (Svindland 2013, S 107) verstanden und bezieht sich auf eine dem Modell gegenüber externe Wahrscheinlichkeit, die reflexionslogisch betrachtet offensichtlich nicht in der modellinternen enthalten sein kann, zumal sie nur fiktiv bestimmbar ist.

In differenzierender Betrachtung verschiedene Wahrscheinlichkeiten als relevant anzuerkennen, heißt davon auszugehen, dass sich bei der Modellierung von zeitsensiblen Phänomenen, die ›Realität des Modells‹ mit fortlaufender Zeit womöglich immer weiter von der ›Realität der Realität‹ entfernt, selbst wenn es mit ›Echtzeitdaten‹ aktualisiert wird. Denn es gilt: »*Ereignisse, über deren mögliches Auftreten wir uns heute nicht im Klaren sind, können bei der Anwendung der Modelle niemals richtig berücksichtigt werden*« (ebd., S. 105, Hervorh. i.O.). So gesehen können im Modell nicht enthaltene Eigenschaften der Realität also erst zu einem späteren Zeitpunkt, dann aber eben auf entscheidende Weise relevant werden.

Grundsätzlich ist der Zusammenhang zwischen Modellierung und modellierter Wirklichkeit daher als kontingent zu verstehen. Belastbare Indizien für die tatsäch-

liche Validität des Modells zeigen sich häufig nur rückblickend, was (›vorblickend‹) die Rede vom »Modellrisiko« plausibel macht.

Es zeigt sich in diesen Überlegungen, wie voraussetzungsreich die Implementierung stochastischer Logik in realen technischen Anwendungen ist. Unwägbarkeiten vieler ML-Anwendungen werden mittels einer solchen Sichtweise zunächst weiter verkompliziert, indem sie als elementare Bestandteile der sozialen Disposition, für die ML sorgt, sichtbar gemacht werden. Diese Perspektive eröffnet vor allem wissenssoziologische Fragestellungen, geht es letztlich doch grundlegendend um den epistemischen Status von ML-Outputs (als Beiträge zu Kommunikation und Wissen.

Derartige epistemologische Reflexionen führen gleichwohl über ML im engeren Sinne hinaus, gelten sie doch für jede modellbasierte Form rechentechnischer Kognition. Schon die Grundannahme, dass die Wirklichkeit mittels probabilistischer Kalküle beschrieben werden kann, erweist sich als voraussetzungsreich – und in historischer Betrachtung als kontingente Errungenschaft. Der britische Wissenschaftsphilosoph und -historiker Ian Hacking widmete sich ausführlich der Emergenz von Wahrscheinlichkeit als epistemischem Phänomen im 19. Jahrhundert. In seiner Geschichte der »Zähmung des Zufalls« (Hacking 2023 [1990]) beschreibt er den Prozess der Erosion deterministischen Denkens im Zuge der Herausbildung einer Wissensordnung, in der statt fixierten Auffassungen (etwa ›wahrer menschlicher Natur‹) fortan Vorstellungen von verteilten Ausprägungen (Menschlichkeit als ›Spektrum diverser, als menschlich betrachteter Eigenschaften‹) prägend wurden. Die Entdeckung der Regelmäßigkeit von Merkmalsverteilungen in sozialen Phänomenen sorgte für Faszination – und bereitete den Weg für langfristige Folgen. Sie wurde ermöglicht durch die Verfügbarkeit von ab dem frühen 19. Jahrhunderten zunehmend offiziell geführten Sozialstatistiken. Zu deren Erstellung und Verwaltung wurden im nachnapoleonischen Europa administrative Institutionen aufgebaut und mit der Aufgabe betraut, Statistiken über eine Vielzahl sozialer Phänomene zu führen und zu veröffentlichen (vgl. Hacking 1990, Kap. 4). Derartige Aufstellungen enthielten nicht selten vor allem Angaben zu Phänomenen der Devianz: Krankheiten, Verbrechen wie auch – nicht zuletzt für die Konstitution der akademischen Soziologie in Frankreich von zentraler Bedeutung[73] – Suizide wurden in ihrem Auftreten zahlenmäßig erfasst und damit für statistische Analysen verfügbar gemacht (vgl. ebd., Kap. 6–8).

Hackings Studie ist hier relevant, weil sie verdeutlicht, dass die für die epistemologischen Implikation von ML maßgeblichen ›Gesetze der Statistik‹ nicht etwa vom Himmel fielen, sondern zunächst vielmehr ›in die Daten gelesen‹ als ›den Daten abgelesen‹ sowie öffentlich legitimiert und verteidigt werden mussten, bevor

73 Vgl. ebd., Kap. 20 sowie klassisch (und auch von Hacking zitiert) Durkheim 1983 [1897].

Zusammenhänge wie die Normalverteilung (Carl Friedrich Gauss, Pierre Simon Laplace, vgl. ebd., Kap. 13 & 21) oder ›Das Gesetz der großen Zahl‹ (Siméon Denis Poisson, vgl. ebd., Kap. 12) in den epistemischen Rang relativ unangefochtener Universalgesetze gelangen konnten. Diese Zusammenhänge sind unbedingt erwähnenswert, weil sie wesentliche Elemente einer epistemischen Ordnung sind, die bis in die Gegenwart wirkt und als zentrale epistemische Stützpfeiler der Konzeptualisierung wie auch der Legitimierung von ML dienen.

Es lässt sich vermittelt über diesen Exkurs eine historische Linie zeichnen, an deren aktuellem Ende Maschinen nicht nur die Vorhersagen zukünftigen menschlichen Kaufverhaltens zu- und anvertraut wird, sondern gesellschaftliche Entwicklungen auf fundamentale Weise von Phänomenen wie dem Klimawandel geprägt werden, die ohne rechentechnische Modellierungsverfahren gar nicht vorstellbar wären.[74] Schon deshalb erscheint es naheliegend, für ein besseres Verständnis technisierter Wahrscheinlichkeitsbestimmungen einen Seitenblick auf Diskurse des 19. Jahrhunderts vorzunehmen. Dies gilt nicht allein, weil die weiteren Entwicklungen der Sozialphysik und Anthropometrie (Adolphe Quetelet) sowie der Eugenik (Francis Galton) rückblickend als Vorboten dunkler ideologischer Gefilde erscheinen müssen, wobei der Zusammenhang von Sozialstatistik, Bevölkerungskontrolle, Rassismus und (nicht nur epistemischer) Gewalt wohl kaum an Relevanz eingebüßt hat[75] und Verknüpfungen von gegenwärtigen Technikentwicklungen mit Ereignissen im 19. Jahrhundert auf überzeugende Weise eine in mancher Hinsicht problematische Vorgeschichte Künstlicher Intelligenz aufzeigen können (vgl. etwa Chun 2021, Amaro 2022).

Das wesentliche Anliegen dieser Studie hingegen ist in erster Linie eine sozialtheoretisch fundierte soziologische Erschließung von ML, für die historische Hintergründe der verschiedenen Phänomenaspekte vor allem in ihrem systematischen Gehalt interessant sind. Bezogen auf dieses Anliegen erlaubt der vorangehende Exkurs ein tiefenschärferes Verständnis des weiter zu erschließenden epistemischen Paradigmas, für das ML exemplarisch steht. Angesichts des probabilistischen Charakters der Outputs algorithmischer Modellierungen sozialer Phänomene birgt der Rückblick in eine Zeit, in der deterministische Weltvorstellungen von indeterministisch-probabilistischen abgelöst wurden, in mancher Hinsicht relativierende Beobachtungen hinsichtlich der gegenwärtig reklamierten epistemischen Schockwirkung von ML (vgl. Roberge/Castelle 2021, S. 2). Die historische Distanz zu den beschriebenen Phänomenen vermag zudem mit größerer Klarheit die Voraussetzungen der ML inhärenten Epistemologie vor Augen führen und im Umkehrschluss

74 In diesem Sinne versteht Benjamin Bratton den Klimawandel als »epistemological accomplishment of planetary-scale computation« (Bratton 2019, S. 10).

75 Siehe dazu weiterführend auch Jürgen Links Theorie des Normalismus in vergangenen und gegenwärtigen Gesellschaften (vgl. Link 1996).

gleichsam zur Offenlegung des kontingenten Charakters gegenwärtiger Annahmen bezüglich der technischen Verarbeitung und Generierung von Wissen beitragen.

Aus einer solchen Perspektive heraus lassen sich einerseits in gegenwärtigen sozialen Kontexten wirksame Prämissen, in denen etwa die Kreditwürdigkeit von Personen automatisch von ML-basierten Programmen bewertet wird, identifizieren und problematisieren. Was heißt es etwa über Kreditnehmer:innen, wenn das eingesetzte ML-Modell auf Basis der im Antrag angegebenen Daten anhand von Vergleichsdaten vorhersagt, dass die Wahrscheinlichkeit eines Zahlungsausfalls 30 Prozent beträgt? Was wird es – vorausblickend auf eine meiner Fallstudien – heißen, dass ein automatisch erzeugtes Bild mit einer Wahrscheinlichkeit von 93 Prozent dem ursprünglichen Bilddatensatz angehören könnte? Und was wird es im Kontext meiner zweiten Fallstudie heißen, dass der Aufenthaltsort einer Honigbiene in einem Beobachtungsstock zu einem beliebigen Zeitpunkt mit einer Wahrscheinlichkeit von 80 Prozent dem Aufenthaltsort zum gleichen Zeitpunkt am nächsten Tag entsprechen wird?

In den vorangehenden Überlegungen zu Korrelation und Kausalität in technischen Arrangements habe ich zu vermitteln versucht, dass sich die bereits erarbeiteten Kerncharakteristika ML als sozialem Phänomen soziologisch nachvollziehbar machen lassen mit einem Verständnis von ML als rechentechnischem Kognitionsverfahren, dessen transformative Effekte vor allem epistemischer Art sind. Als zentral hat sich die stochastische Operationslogik algorithmischer Modelle erwiesen, mithilfe derer sich probabilistische Beschreibungen von Datenbeziehungen oder mit gewisser Wahrscheinlichkeit zum Modelldatensatz passende synthetische Outputs erzeugen lassen. Historische Perspektiven auf die Verknüpfung von Rechentechnik, Information und Kommunikation (Weaver) und die »Zähmung des Zufalls« (Hacking 2023 [1990]) im 19. Jahrhundert haben dabei geholfen, Möglichkeiten der soziologischen Perspektivierung ML zu identifizieren, d.h. Problematisierungsweisen von ML als sozialem Phänomen, die für die im nächsten Kapitel anschließende sozialtheoretische Konzeptualisierung und für den gesamten weiteren Verlauf der Untersuchung maßgeblich sein werden. Als letzter Aspekt der die Theoriebildung vorbereitenden Erörterung von ML sollen nun zunächst jedoch noch dessen zeitliche und räumliche Extensionen erschlossen werden.

Zeitliche und räumliche Dimensionen

Unabhängig von der Beurteilung der technikheoretischen und epistemologischen Fragen, wie sie sich im vorangehenden Abschnitt gestellt haben, kann nach den Ermöglichungsbedingungen der Umstände gefragt werden, die diesen Fragen überhaupt erst Plausibilität verleihen. Schon in oberflächlicher Betrachtung zeigt sich, dass ML für die Entfaltung seines Wirkens als zeitlich differenzierter Vorgang in vie-

lerlei Hinsicht auf Infrastrukturen angewiesen ist, die seinem praktischen Einsatz in Form technischer Anwendungen vorausgehen.

Gewissermaßen am offensichtlichsten zeigt sich dies in materiellen Abhängigkeiten, etwa von energetischen Ressourcen.[76] Offenkundig wurde der Aufstieg von ML in den letzten zehn Jahren durch die Verfügbarkeit zunehmend wachsender Rechenleistung zur zeitlich effizienten algorithmischen Verarbeitung umfangreicher Datensätze begünstigt (und überhaupt erst möglich) – und nicht eben etwa primär durch die Entwicklung neuer, ›besserer‹ Algorithmen, auch wenn die in der Einleitung geschilderten Entwicklungen schließlich ebenso weiterführende Forschungen an Algorithmen stimuliert haben (vgl. Sudmann 2018, S. 22f.). Damit geht einher, dass parallel zum gegenwärtigen KI-Hype auch die Nachfrage an für die Konstruktion von Rechentechnik relevante Rohstoffe – etwa Silizium oder Germanium als Halbleiter – weiter steigt, auch wenn diese wachsende Abhängigkeit von nicht unendlich verfügbaren natürlichen Ressourcen keineswegs allein der jüngeren KI/ML-Entwicklung, sondern der Gänze all jener umfassenden Computerisierungs- bzw. Digitalisierungsprozessen mindestens der letzten fünfzig Jahre zuzuschreiben ist.[77]

Darüber hinaus ist ML wesentlich auf die Verfügbarkeit von Daten angewiesen. Das *Lernen* der Maschine meint immer *Lernen aus Daten*. Algorithmische Modelle sind ohne geeignete Datensätze zur Modellierung interessierender Zusammenhängen schlichtweg nicht denkbar.[78] Die Verfügbarkeit von Daten für einen gegebenen Zusammenhang in geeigneter Form ist allerdings alles andere als voraussetzungslos. Sie impliziert Anforderungen an das Verfügbar*machen*, aber auch das Verfügbar*halten* von Daten. Neben Hardware wie adäquaten Speichermedien und Übertragungskanälen erweist sich die bereits erwähnte »ghost work« (Suri/Gray 2019) in vielen Zusammenhängen als unverzichtbare Infrastruktur von ML: ohne die Erledi-

76 Der hohe Energiebedarf der Trainingsverfahren mancher Modelle hat sich mit Blick auf Umweltfolgen als einschlägiger Aspekt der Kritik an ML etabliert, vgl. etwa Whittaker 2021, Bender et al. 2021.

77 Was die geopolitische Brisanz der Beziehung zwischen Taiwan und China begründet, ist *Taiwan Semiconductor Manufacturing Company Limited* (TSMC) doch der weltgrößte Hersteller von Halbleitern.

78 Vgl. weiterführend zum Zusammenhang von Lernen und Datenverarbeitung Parisi 2018. Dieser Aspekt impliziert übrigens auch, dass vom Modell ›Gelerntes‹ praktisch der Gegenwart immerzu ›hinterherhängen‹ muss. Zwar geschieht die konkrete Modellanwendung in einer konkreten Gegenwart, doch kann sich ML nicht unmittelbar auf diese beziehen, weil jede Modellierung auf Basis von Daten stattfindet, die in der Vergangenheit erhoben worden sind, auch wenn Modellierungen adaptiv sein können. Es ist also gerade das Gegenteil des aus der Psychologie bekannten »recency bias« anzunehmen – zumindest so lange ML nicht zur unmittelbaren Einspeisung von Echtzeitdaten in die Modellierung fähig ist.

gung der für die Erhebung, Aufbereitung und Pflege von Datensätzen anfallenden manuellen Tätigkeiten wäre ML in seinen gegenwärtigen Gestalten nicht denkbar.

Neben der Bedingung, *dass* die für ML-Modelle benötigten Daten verwertbar sind, zeigt sich, dass ML mit Blick darauf, *was* die Daten enthalten, als ein Instrument der »Wissensextraktion« (Joler/Pasquinelli 2020, Übers. RG) beschrieben werden kann. So erscheint ML gesellschaftstheoretisch in (post-)marxistischer Betrachtung nicht etwa als Imitation menschlicher oder biologischer Intelligenz, sondern in erster Linie als akkumulierte Arbeit (vgl. Pasquinelli 2023). Das Datensätzen automatisch extrahierte Wissen mitsamt der darin enthaltenen Potenziale für intelligente Beobachtungen verweist in einer solchen Sichtweise auf die Arbeit, die es brauchte, um die verwendeten Daten überhaupt zu erzeugen und verfügbar zu machen. Demnach wäre zu betonen, dass die manuelle Annotationstätigkeiten im Kategorisieren von Bilddaten – man denke an für die Forschung so wesentliche Datensätze wie *ImageNet* – die anschließend algorithmisch zu modellierenden Inhalte der Datensätze überhaupt erst hervorgebracht hat. Was in einem späteren Stadium als ›intelligente‹ *Computer Vision* gefeiert werden kann, wäre jedenfalls schlichtweg undenkbar, ohne dass vorab das korrekte manuelle Verknüpfen von Bildern mit entsprechenden Tags erfolgt wäre. Ein Datensatz, dessen *ground truth* gemessen an der Wirklichkeit hingegen fehlerhaft ist, entzieht ›maschineller Intelligenz‹ ihre materielle Grundlage.[79] In diesem Zusammenhang stellt sich daher die sozialtheoretische brisante Frage, welche Relevanz der automatischen Produktion von (nicht-trivialen) Analysen des Datensatzes im Sinne des Sichtbarmachens seiner Eigenschaften zuzurechnen ist.[80]

Ähnliches gilt für generative Verfahren und ob Prädiktionen eines Modells in erster Linie als Derivate oder Extra- bzw. Interpolationen der Datenbasis gelten soll-

79 Eine nicht zu unterschätzende Fehlerquelle für ML-Anwendungen, vgl. Jaton 2024, Jaton 2021, Kap. 1 & 2.

80 Entlang solcher Überlegungen stellt sich die sozial- wie auch gesellschaftstheoretisch hochrelevante Frage, wie darüber zu entscheiden ist, unter welchen Bedingungen Wissen ›bloß extrahiert‹ und wo es dementgegen ›aktiv erzeugt‹ wird. Hieran schließt weiter unten explizit das sozialtheoretische Problem der Zurechnung von Potenzialen und Fähigkeiten (i.S.v. Agency oder Kommunikationsfähigkeit) an, mit dem die Zuschreibung eines bestimmten sozialen Status (Akteur vs. Technik) in engem Zusammenhang steht. Darüber hinaus halte ich die semantische Verwandtschaft von ›*Zu*rechnung‹ (etwa einer Handlung) und *be*rechnender Technik (*computational technology*) für interessant, führt diese doch zu interessanten Fragen mit Blick auf Adressierbarkeit. So wie rechentechnische Prozesse quasi-fraglos immer als *Be*rechnungen von etwas und durch etwas identifizierbar anmuten, erscheint die Verknüpfung von Adressierbarkeit und Rechentechnik, wie Ranjodh Singh Dhaliwal in seiner Kontemplation der Frage »What Even Is Computation?« bemerkt, in kulturgeschichtlicher Reflexion der Möglichkeitsbedingungen von Adressierbarkeit (etwa von Staatsbürger:innen qua administrativer Fixierung von Wohn*adressen*) womöglich doch kontingenter, als erste Eindrücke suggerieren mögen (vgl. Dhaliwal 2022).

ten, ist eine Frage, die nicht nur sozialtheoretisch interessant ist (vgl. etwa Bonhasse-Gahot 2017[81]), sondern gegenwärtig auch in urheberrechtlicher Hinsicht intensiv diskutiert wird.[82] Jedenfalls erweist sich an diesem strittigen Punkt zweifellos, *dass* ML-Anwendungen jedweder Couleur auf die Verfügbarkeit geeigneter Datensätze angewiesen sind. Dass Datensätze eigene Entstehungsgeschichten zu eigen haben, muss diesen nicht offenkundig entnehmbar sein, gibt der soziologischen Analyse von ML jedoch einen Hinweis auf die Vielfalt an potenziell instruktiven Assoziationen, die von einem Softwareartefakt (als Kontrolloberfläche einer ML-Anwendung) aus verfolgt werden können.

Dieser Umstand, zeigt, dass Operationen von ML-Modellen – ob im Vorgang des Trainings oder in ihrer Anwendung – in ihrer sozialen Materialität zudem eine distribuierte genetische Signatur tragen, die auf eine Vielzahl von Materialien und Praktiken an verschiedenen Orten und unter Beteiligung verschiedener Entitäten verweist. Derartigen infrastrukturellen Signaturen folgend lässt sich ein vielfältiges Netzwerk von Interdependenzbeziehungen erschließen, dessen Analyse die Möglichkeitsbedingungen maschinellen Lernens zu offenbaren vermag. Einem solchen Anspruch folgt etwa Kate Crawfords »Atlas of AI« (Crawford 2021). Die Autorin führt in der Legende ihres Atlas – um im Bild zu bleiben – eine Vielzahl sozialer Phänomene auf, aus deren Zusammenwirken sich KI/ML gegenwärtig speist. Zu diesen zählt Crawford neben bereits Erwähntem wie natürliche Ressourcen und menschlich verrichtete Arbeit ebenso Klassifikationssysteme und logistische Arrangements, aber auch (staatlich verfasste) politische Ordnungen und Herrschaftsverhältnisse im weiteren Sinne (vgl. ebd., S. 8). Crawfords Untersuchung schafft einen Eindruck der Größenordnung und Heterogenität von für ML erforderlichen Mobilisierungen, die verschiedene Extraktions- und Ausbeutungsverhältnisse ebenso einschließen wie seltene Erden und Datenbanken sowie menschliche Arbeitskraft. Die sozialen Folgen der Entstehung und Verbreitung ›Künstlicher Intelligenz‹ gehen anders gesagt – und manchen Mythen entgegen[83] – also wesentlich auf soziale Phä-

81 Bezugspunkt der Beurteilung sind dabei die Trainingsdaten – deren »convex hull« bzw. die Grenzen des multidimensionalen Vektorenraums, den diese beschreiben – und damit die Frage, wie sich Outputs zu diesen Verhalten. Für ein Beispiel einer Position *pro* Extrapolation vgl. etwa Balestriero et al. 2021.

82 Angesichts der zunehmenden Verbreitung etwa von Bildgenerierungsverfahren wie Midjourney – als visuellem Äquivalent von LLMs – wäre etwa zu klären, ob bildnerisch tätige Künstler:innen und Grafiker:innen, deren Werke im Modelldatensatz vorhanden sind, einen Anspruch auf Tantiemen haben sollten, wenn User:innen aus Prompts wie »Eine Grafik im Stile von […]« Bilder generieren und dabei für die Nutzung des Dienstes Zahlungen etwa an Midjourney leisten oder gar selbst für kommerzielle Zwecke nutzen, vgl. Escalante-De Mattei 2023.

83 Seit der Antike bestehen Mythen über Wesen und Potenziale von Maschinen, die auf dem Weg in die moderne Gesellschaft (etwa La Mettrie) verschiedene Metamorphosen durchlaufen haben. Der Computer steht indes für eine Zäsur in der Imagination der Maschine, indem

nomene zurück, die weniger verheißungsvoll von utopischen Zukünften künden, als dass sie die Realität einer global operierenden extraktiven Technikindustrie vor Augen führen. Angesichts dieser Umstände, so stellt Crawford (in fragwürdiger Zuspitzung) fest, sei KI/ML entgegen manchen Verlautbarungen »neither artificial nor intelligent«, sondern

> an idea, an infrastructure, an industry, a form of exercising power, and a way of seeing; it's also a manifestation of highly organized capital backed by vast systems of extraction and logistics, with supply chains that wrap around the entire planet. (Crawford 2021, S. 18f.)

Wenngleich ich Crawfords Identifizierung von KI mit einer solchen Vielzahl in sich komplexer Phänomene sozialtheoretisch nicht folgen kann, und die Negativformel »neither artificial nor intelligent« allein als rhetorische Figur zur ›Abkühlung‹ von Technikindustriepropaganda einleuchtend finde, halte ich ihre Beobachtungen im Ganzen dennoch für instruktiv. Ihre Beschreibungen der für ML als gesellschaftlichem Phänomen charakteristischen Interdependenzverhältnisse liefern allemal Einsichten über den Grad an Integration, den ML über verschiedene gesellschaftliche Zusammenhänge hinweg auszeichnet. Andererseits lässt sich ihnen eine Vielzahl von Hinweisen auf die heterogenen Folgen des Einsatzes von KI/ML entnehmen. Diese umfassen – so in grober Reformulierung Crawfords – mithin semantische, industrielle, politische ebenso wie epistemologische, ökonomische, logistische und ökologische Aspekte. Statt ML jedoch mit den Phänomenen selbst gleichzusetzen, was ich aufgrund der resultierenden Unterschiedslosigkeit analytisch nicht zielführend finde, ließe sich doch für jeden einzelnen der genannten Zusammenhänge gewissermaßen die soziale (Binnen-)Wirkung des Einsatzes von ML untersuchen. Für einen Beurteilung von deren Ausmaß – so sollte das bisherige Argument gezeigt haben – ist das Zusammenwirken einer Vielzahl von Entitäten in Betracht zu ziehen.

Drei Phasen

Neben dieser ökologischen Perspektive, die Wechselbeziehungen mit anderen sozialen Zusammenhängen in das Blickfeld bringt, führt die Betrachtung der zeitlichen Dimension von ML im Anschluss an die vorgetragenen Überlegungen an dessen infrastrukturellen Voraussetzungen zu einer in Phasen differenzierenden Heuristik. In seiner Abhängigkeit von Infrastrukturen erscheint ML weniger als radikale

---

er die Repräsentation von Universalismus ermöglicht; eine Position, die vorher göttlichen Wesen vorbehalten schienen, vgl. Sudmann 2018a, S. 18.

Transformation – jedenfalls in keiner der drei Phasen für sich genommen. Stattdessen erweist es sich soziologisch als ein auf verschiedene Phasen verteilter fortwährender Akkumulationsvorgang von Daten, Einstellungen, Geräten und weiteren beteiligten Elementen (vgl. Mackenzie 2017, S. 5f.). Diese werden durch eine Reihe von sozialen Praktiken als Ressourcen mobilisiert. Schematisch betrachtet vollzieht sich ML als Akkumulation über die Phasen der *Datenerhebung*, *Datenmodellierung* und *Modellanwendung* hinweg. Für sich genommen verweiset jede der drei Phasen auf je distinkte Praktiken. Als arbeitsteiliger Prozess ähnelt ML damit in gewisser Hinsicht dem für den Industriekapitalismus typischen Fließbandprinzip (vgl. Pasquinelli/Joler 2020, S. 1264f.).

In funktionaler Hinsicht entspricht diese Einteilung Florian Jatons Vorschlag der Schematisierung verschiedener »courses of action« (vgl. Jaton 2021, S. 23f.) hin zur Anwendung eines Algorithmus: *Ground Truthing*, *Programming* und *Formulating* (vgl. ebd.). Der »Data Collection«-Phase bei Pasquinelli und Joler entspricht dabei das schon erwähnte »Ground Truthing« (vgl. ebd., Kap. 1), das die Definition einer Datenbasis für das Modell leistet. Dieser Vorgang legt den Wirklichkeitsbezug des Modells fest: einem interessierenden Wirklichkeitsphänomen wird eine diese möglichst akkurat abbildende Datenbasis zugeordnet. Damit diese zur ›Basiswahrheit‹ des Modells werden kann, muss sie zunächst alle Daten in eine verarbeitbare Form bringen, die zuverlässigen Zugriff auf diese ermöglicht. Ist dieser Schritt erfolgt, kann die Programmierung des Modells beginnen; diesem Schritt der Datenmodellierung bei Pasquinelli/Joler entsprechen bei Jaton zwei distinkte »courses of action«: »Programming« und »Formulating«. Die Unterscheidung entspringt einer unmittelbar nachvollziehbaren praktischen Beobachtung: Die verbindliche mathematische Formulierung eines Modells setzt in aller Regel bereits verschiedene Versuche der Programmierung eines adäquaten Modells voraus.

Jatons ethnografische Studie, die von teilnehmenden Beobachtungen der Praxis in einem mathematisch-informatischen Labor ausgeht, macht sich explizit Algorithmen zum Untersuchungsgegenstand und verfolgt präzise, was in deren ›Konstitution‹ einfließt. Gegenüber dem Modell des *Nooscope*[84] bei Pasquinelli und Joler bedeutet dies zugleich eine weitere wie engere Perspektive: weiter in dem Sinne, dass Jaton Algorithmen in einem allgemeineren Sinne untersucht, wobei solche Algorithmen, die im maschinellen Lernen genutzt werden, gesondert diskutiert werden als ein besonderer Typus, den auszeichnet, dass die Formulierung des Algorithmus automatisiert wird (vgl. Jaton 2021, S. 266–281). Enger ist Jatons Perspektive hingegen in dem Sinne zu verstehen, dass die Phase der für Pasquinelli und Joler wesentlichen Modellanwendung nicht vorkommt, da diese nicht mehr in den Ge-

84 Ein dem Griechischen entlehntes Komposit, das *skopein* (sehen, betrachten) und *noos* (Wissen) kombiniert.

genstandsbereich der *Konstitution* eines Algorithmus fällt, sondern eben in dessen *Anwendung*.

Dieser Bereich ist jedoch für das Erkenntnisinteresse der vorliegenden Untersuchung zentral, da ML hier wesentliche Aspekte seiner sozialen Wirkung entfaltet. So instruktiv Jatons Untersuchung für ein angemessenes Verständnis der Genese von Algorithmen ist, erweist es sich für eine sozialtheoretisches Auseinandersetzung doch als unerlässlich, an diese Genese anschließende »courses of action« auf dem Weg zur ›Konstitution‹ von maschinellem Lernen als Praxis ebenso mit in den Blick zu nehmen. Nur mithilfe einer solchen integrierten Perspektive kann es gelingen, ML aus der Konstitution seiner Elemente – mitsamt der potenziellen Effekte dieses Prozesses auf alles Folgende – wie auch aus seiner Realisierung in praktischer Anwendung heraus zu beschreiben. Dies gilt schon vor dem Hintergrund der basalen techniksoziologischen Einsicht, dass technische Objekte in ihrer konkreten Nutzung in aller Regel maßgeblich von der ursprünglich durch deren Design intendierten Nutzung abweichen (vgl. Bijker/Law 1992), als unbedingt geboten im Sinne eines differenzierten Verständnisses von ML als sozialem Phänomen.[85] Es ist wesentliches Anliegen des empirischen Teils dieser Studie, die Dynamiken praktischer Anwendungen von ML in ebendieser Vielzahl von Hinsichten zu erfassen. Dass sich mittels einer solchen Perspektivierung des Untersuchungsgegenstandes gleichsam epistemologische wie gesellschaftstheoretisch instruktive Einsichten finden lassen können, zeigt Wendy Chun Studie mit dem Titel »Discriminating Data« (Chun 2021). Dort leistet die Autorin gewissermaßen eine Übersetzung von vier technischen in sozialen Funktionen von ML: *Korrelation (1)* sei zunächst die Paarung von Datenpunkten als Ausgangspunkt von möglichen Erkenntnissen; *Diskriminierung (2)* separiere dann (oft versteckt, vgl. ebd., S. 96) normativ – spezifischer: zumeist »homophil« (ebd.) operierend: Gleiches zu Gleichem – korrelierte von unkorrelierten Daten; die folgende *Authentifizierung* (3) lege auf Basis der Korrelate positiv Gruppen fest, die schließlich *Anerkennung (4)* bzw. Identifikation erfahre, bevor das derart mittels ML erzeugte Wissen wirksam werden kann. Dies gehe aufgrund des konservativen Vergangenheits-Bias der Daten zumeist mit Normalisierungseffekten einher:

> When correlation works, it does so by making the present and future coincide with a highly curated past. (Chun 2021, S. 52)

85 Eine Einsicht übrigens, die sozialtheoretisch auch auf andere zeitsensible Zusammenhänge übertragbar ist. In Arnold Gehlens Handlungs- und Institutionentheorie etwa wird dieser Nexus als die Ausbildung von Selbstzwecken beschrieben, die handlungstheoretisch auf der Möglichkeit der Trennung von Motiv und Zweck beruht. Institutionen fußen in ihrer Konstitution auf solchen Vorgängen des Umschlagens, die Handlungen »sekundäre Zweckmäßigkeiten« verleihen, welche wiederum später zu institutionellen Eigenwerten werden können, vgl. Gehlen 2022 [1956], Kap. 7–9 sowie Rehberg & Groß 2022, S. 503ff.

Chuns Argument konzentriert sich wesentlich auf machttheoretische Aspekte der Ausbreitung eines epistemischen Dispositivs dessen Wurzeln sie bis in die Hochzeit der (britischen) Eugenik im 19. Jahrhunderts zurückverfolgt. Aus jener Bewegung sei nicht nur Rassismus im Deckmantel der Wissenschaft hervorgegangen, sondern sei diese ebenso für wesentliche Grundlagen von Methoden etwa der statistischen Regression und Korrelation verantwortlich – wobei wohl gerade die Gleichzeitigkeit beider Entwicklungen zu denken gebe.[86] Wenngleich machttheoretische Überlegungen an dieser Stelle nicht unmittelbar relevant sind, erweist sich Chuns Beschreibung der für die Entfaltung von ML notwendigen Schritte zur Etablierung epistemischer Normen allemal als instruktiv, weil damit eine differenzierende Sicht auf ML möglich wird. Im Zusammenhang mit den anderen beiden eben referierten Schemata gewinnt im hier sich entwickelnden Verständnis von ML als facettenreichem Phänomen zunehmend Konturen.

ML erscheint – so ein erstes Zwischenfazit am Ende dieses Kapitels – soziologisch als rechentechnisches Arrangement, das auf Grundlage von Datensätzen mittels algorithmischer Modellierung in nicht-trivialer Weise Outputs für gegebene Inputs errechnet. In praktischen Anwendungen werden dieses als Beobachtungen der Zusammenhänge beobachtet, die als im Datensatz abgebildet angenommen werden. ML entfaltet mithin in Form analytischer oder generativer Outputs seine soziale Wirkung als Technik. Deren Produktion setzt verschiedene infrastrukturelle Gegebenheiten voraus und lässt ML als zeitlich in drei distinkte Phasen differenzierbares Phänomen erscheinen. Einer ersten Phase der Auswahl und Aufbereitung eines Datensatzes (ggf. inkl. Datenerhebung) als *Ground Truth* des Anwendung folgt die Trainingsphase, in der die algorithmische Selbstprogrammierung des Modells stattfindet, bevor schließlich in der Anwendungsphase das Modell als Maschine eingesetzt wird, die für gegebene Inputs Outputs produziert.

Vermittelt über die Beobachtung der Outputs wie der Operationslogik algorithmischer Modellierungen offenbart ML eine ambivalente soziale Verfasstheit. Beobachter:innen von ML werden gleichermaßen mit vertrauten wie befremdlichen Aspekten konfrontiert, wodurch es sich als soziales Phänomen eindeutigen Bestimmungen entzieht. Als maßgebliche Ursache dafür wurde infolge der Verdatung von sinnförmig verfassten Beobachtungen und Wissen, der stochastischen Modellierung der in diesem Zug erzeugten Datensätze (›Lernen‹) sowie der Rückübersetzung der maschinell produzierten Outputs in sinnförmige Beobachtungen durch Beobachter:innen eine maßgebliche epistemische Lücke identifiziert. Der probabilistische Charakter von Modelloutputs verunmöglicht ein Verständnis von ML als ›klassische‹, im Medium der Kausalität operierende Technik und mithin die Rekonstruktion der rechentechnischen Kognitionsprozesse in ML als Sinnprozesse.

86 Vgl. dazu auch die Ausführungen zu Hackings »Zähmung des Zufalls« (S. 82f.).

Folglich ist ML als soziales Phänomen vor allem durch die Schwierigkeiten eindeutiger Bestimmbarkeit geprägt, weshalb jede ML-Anwendung prinzipiell – trotz vollständig bestimmter Outputs – in potenziell wesentlicher Hinsicht unbestimmt (und womöglich unbestimmbar) erscheinen muss. Für diese These kann eine Reihe von Gründen angeführt werden, so etwa erstens die Opazität nicht-trivialer algorithmischen Modellierung mitsamt daraus folgenden Schwierigkeiten des Verstehens von Modelloutputs für deren Beobachter:innen, weshalb zweitens auch der epistemische Status von Modelloutputs fragwürdig erscheinen muss und mithin drittens gleichsam der Status von ML als an sozialen Praktiken beteiligte rechentechnische Entität überhaupt.

Die vorliegende Untersuchung wird im weiteren Verlauf, den erarbeiteten Einsichten und Fragenstellungen folgend, in Form zweier ethnografischer Studien in Kunst und Wissenschaft empirisches Material zusammentragen, um fallweise die sozialen Voraussetzungen praktischer Manifestationen von ML zu untersuchen. Ebenso sollen dabei die Umstände, Bedingungen und Dynamiken des praktischen Einsätze von ML in den Blick genommen werden. Insbesondere wird die empirische Untersuchung, wie weiter unten ausführlicher zu zeigen sein wird, das Motiv der epistemischen Schockwirkung von ML aufnehmen und sich deshalb auf die praktischen Probleme mit ML sowie deren Bearbeitung in der Absicht auf produktive Wendungen fokussieren.

Diagnosen wie jener eines epistemischen Schocks im Zusammenhang mit der Verbreitung von ML und die anhaltenden Kontroversen um dessen Beurteilung verweisen nicht nur auf die gesellschaftliche Relevanz von ML, sondern unterstreichen vor allem auch die Relevanz und das Potenzial sozialtheoretischer Auseinandersetzung im gegenwärtigen Ringen um eine adäquates Verständnis von ML und KI im weiteren Sinne. Sein kognitiver Charakter lässt ML für den Einsatz im Zusammenhang mit der Verarbeitung und Produktion von Wissen geeignet erscheinen und schon der Umstand, *dass* sich praktisch entsprechende Erwartungen bilden, die zu praktischen Versuchen der Beteiligungen von ML an Kommunikation führen, sorgt für soziale Folgen, die unbedingt untersuchungswürdig sind.

Das folgende Kapitel ist im Anschluss an die bereits entwickelten Überlegungen zunächst einer vorläufigen sozialtheoretischen Konzeptualisierung von ML gewidmet. Danach werde ich ausgehend von meinem konzeptuellen Rahmen einen Vorschlag zur empirischen Operationalisierung dieses Verständnisses von ML erarbeiten, und diesen mit methodologischen und methodischen Überlegungen zur angemessenen empirischen Untersuchung von ML verknüpfen. Das bereits entwickelte Motiv der Bestimmbarkeitsschwierigkeiten von ML wird darin eine methodologische Wendung erfahren, infolge derer die empirische Analyse von ML als situative Problemanalyse plausibilisiert wird. Bereits an dieser frühen Stelle der Untersuchung zeigt sich, dass die charakteristische Vielzahl von Unwägbarkeiten in der Betrachtung von ML bestenfalls produktiv in die empirische Untersuchung anleitende

Fragen transformiert werden sollte. Mit diesem Vorgehen folge ich der Vermutung, dass sich für die Spannungen und Ambivalenzen, wie sie die vorhandenen soziologischen Beschreibungen von ML kennzeichnen, auf die eine oder andere Weise Resonanzen in den praktischen Umsetzungen von ML finden lassen sollten.